

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 24, 2004, 14:54:57 ; Search time 61.1194 Seconds  
(without alignments)  
69.343 Million cell updates/sec

Title: US-09-641-801-34  
Perfect score: 89  
Sequence: 1 MHQPPQQLPPTVMFP 15

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*  
1: geneseqp1980s:\*  
2: geneseqp1990s:\*  
3: geneseqp2000s:\*  
4: geneseqp2001s:\*  
5: geneseqp2002s:\*  
6: geneseqp2003as:\*  
7: geneseqp2003bs:\*  
8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	89	100.0	15	4	AAB72531	Aab72531 Colostrin
2	89	100.0	15	4	AAB59334	Aab59334 Ewe colos
3	89	100.0	15	4	AAB72279	Aab72279 Colostrin
4	89	100.0	15	4	AAB72563	Aab72563 Colostrin
5	89	100.0	15	5	AAO14610	Aao14610 Neural ce
6	89	100.0	15	5	AAM51066	Aam51066 Colostrin
7	89	100.0	15	5	AAE20261	Aae20261 Colostrin
8	81	91.0	222	2	AAW00679	Aaw00679 Beta-case
9	80	89.9	209	2	AAR80281	Aar80281 Methyl or

10	80	89.9	209	2	AAR95609	Aar95609	Bovine be
11	80	89.9	209	2	AAW54021	Aaw54021	Modified
12	74	83.1	209	4	AAB60348	Aab60348	Bovine be
13	57	64.0	10	4	AAB72523	Aab72523	Colostrin
14	57	64.0	10	4	AAB72270	Aab72270	Colostrin
15	57	64.0	10	4	AAB72555	Aab72555	Colostrin
16	57	64.0	10	5	AAO14601	Aao14601	Neural ce
17	57	64.0	10	5	AAM51059	Aam51059	Colostrin
18	57	64.0	10	5	AAE20252	Aae20252	Colostrin
19	53	59.6	60	5	ABP64119	Abp64119	Human ORF
20	53	59.6	352	7	ADC10098	Adc10098	Human NOV
21	53	59.6	357	7	ADC10100	Adc10100	Human NOV
22	53	59.6	359	4	AAM38825	Aam38825	Human pol
23	53	59.6	359	4	AAM38823	Aam38823	Human pol
24	53	59.6	375	4	AAM38822	Aam38822	Human pol
25	53	59.6	376	4	AAM38824	Aam38824	Human pol
26	53	59.6	406	4	AAM40608	Aam40608	Human pol
27	53	59.6	419	4	AAM78414	Aam78414	Human pro
28	53	59.6	425	4	AAM79398	Aam79398	Human pro
29	53	59.6	431	4	AAM40611	Aam40611	Human pol
30	53	59.6	431	4	AAM40610	Aam40610	Human pol
31	53	59.6	431	4	AAM40609	Aam40609	Human pol
32	52	58.4	11	4	AAE07192	Aae07192	Colostrin
33	52	58.4	12	4	AAE07202	Aae07202	Modified
34	52	58.4	1596	4	ABB70845	Abb70845	Drosophil
35	51	57.3	10	4	AAG78781	Aag78781	Stachybot
36	51	57.3	256	4	ABG14800	Abg14800	Novel hum
37	51	57.3	266	4	ABG14799	Abg14799	Novel hum
38	51	57.3	545	7	ADC03404	Adc03404	Rice flow
39	50	56.2	111	4	AAG73513	Aag73513	Human gen
40	50	56.2	134	4	AAG73512	Aag73512	Human gen
41	50	56.2	149	4	AAO05346	Aao05346	Human pol
42	50	56.2	327	7	ADC87451	Adc87451	Human GPC
43	50	56.2	337	6	ABU00184	Abu00184	Human nov
44	50	56.2	751	5	ABG69453	Abg69453	Human nov
45	50	56.2	751	6	ABU60658	Abu60658	Human kin

# ALIGNMENTS

RESULT 1

AAB72531

ID AAB72531 standard; peptide; 15 AA.

XX

AC AAB72531;

XX

DT 09-MAY-2001 (first entry)

XX

DE Colostrinin peptide #32.

XX

KW Dermatological; oxidative stress regulator; colostrinin.

XX

OS Unidentified.

XX

PN WO200112650-A2.

XX

PD 22-FEB-2001.  
 XX  
 PF 17-AUG-2000; 2000WO-US022665.  
 XX  
 PR 17-AUG-1999; 99US-0149310P.  
 XX  
 PA (TEXA ) UNIV TEXAS SYSTEM.  
 XX  
 PI Stanton GJ, Hughes TK, Boldogh I;  
 XX  
 DR WPI; 2001-218342/22.  
 XX  
 PT Modulating oxidative stress level in a cell, involves contacting the cell  
 PT with an oxidative stress regulator selected from colostrinin, its  
 PT constituent peptide, analog or their combinations.  
 XX  
 PS Claim 6; Page 26; 48pp; English.  
 XX  
 CC The present invention relates to a method for modulating the oxidative  
 CC stress level in a cell or a patient, comprising contacting the cell with,  
 CC or administering to the patient, an oxidative stress regulator selected  
 CC from colostrinin, or its constituent peptide (e.g. the present peptide),  
 CC to change the level of an oxidising species in the cell. The method can  
 CC be used to treat oxidative damage to skin, by decreasing or preventing an  
 CC increase in the level of damage to a biomolecule of the patient  
 XX  
 SQ Sequence 15 AA;

Query Match 100.0%; Score 89; DB 4; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 9.2e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHQPPQPLPPTVMFP 15  
 |||||  
 Db 1 MHQPPQPLPPTVMFP 15

# RESULT 2

AAB59334

ID AAB59334 standard; peptide; 15 AA.

XX

AC AAB59334;

XX

DT 21-MAR-2001 (first entry)

XX

DE Ewe colostrinin peptide fragment C-9.

XX

KW Sheep; colostrinin; proline rich polypeptide; colostrum; immune disorder;  
 KW central nervous system disorder; dietary supplement; beta-amyloid plaque.

XX

OS Ovis sp.

XX

PN W0200075173-A2.

XX

PD 14-DEC-2000.

XX

PF 02-JUN-2000; 2000WO-GB002128.

XX  
 PR 02-JUN-1999; 99GB-00012852.  
 XX  
 PA (REGE-) REGEN THERAPEUTICS PLC.  
 XX  
 PI Georgiades JA;  
 XX  
 DR WPI; 2001-071058/08.  
 XX  
 PT Peptides having an N-terminal amino acid sequence isolated from  
 PT colostrinin for treating e.g. disorders of the central nervous system and  
 PT immune system, viral and bacterial infections, and diseases characterized  
 PT by amyloid plaques.  
 XX  
 PS Claim 7; Page 27; 63pp; English.  
 XX  
 CC The present invention provides the sequences of a number of peptides  
 CC found in ewe's colostrinin. Colostrinin is the proline-rich polypeptide  
 CC fragment of colostrum. These peptides can be used in the treatment of  
 CC central nervous system disorders such as senile dementia, Parkinson's  
 CC disease, Alzheimer's disease, psychosis and neurosis, immune system  
 CC disorders such as bacterial and viral infections, to improve the  
 CC development of a child's immune system, as a dietary supplement, and to  
 CC promote the dissolution of beta-amyloid plaques  
 XX  
 SQ Sequence 15 AA;

Query Match 100.0%; Score 89; DB 4; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 9.2e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MHQPPQPLPPTVMFP 15  
 |||||  
 Db 1 MHQPPQPLPPTVMFP 15

RESULT 3  
 AAB72279

ID AAB72279 standard; peptide; 15 AA.  
 XX  
 AC AAB72279;  
 XX  
 DT 14-MAY-2001 (first entry)  
 XX  
 DE Colostrinin derived cytokine inducing peptide SEQ ID 34.  
 XX  
 KW Colostrinin; immune response; cytokine; blood cell proliferation;  
 KW central nervous system disorder; neurological disorder; mental disorder;  
 KW dementia; neurodegenerative disease; Alzheimer's disease; psychosis;  
 KW neurosis; infection.  
 XX  
 OS Synthetic.  
 XX  
 PN WO200111937-A2.  
 XX  
 PD 22-FEB-2001.  
 XX

PF 17-AUG-2000; 2000WO-US022818.  
 XX  
 PR 17-AUG-1999; 99US-0149311P.  
 XX  
 PA (TEXA ) UNIV TEXAS SYSTEM.  
 PA (REGE-) REGEN THERAPEUTICS PLC.  
 XX  
 PI Stanton GJ, Hughes TK, Boldogh I, Georgiades J;  
 XX  
 DR WPI; 2001-202804/20.  
 XX  
 PT Inducing a cytokine and modulating an immune response, useful for  
 PT treating central nervous system diseases and bacterial and viral  
 PT infections, comprises administering colostrinin as an immunological  
 PT regulator.  
 XX  
 PS Claim 1; Page 34; 50pp; English.  
 XX  
 CC Sequences AAB72246 - AAB72275 represent peptides derived from clostrinin,  
 CC a proline rich polypeptide aggregate contained in colostrum. The peptides  
 CC have immune response modulatory activity, and are capable of inducing  
 CC cytokines. Colostrinin and its derived peptides are useful for inducing  
 CC cytokine production, for modulating an immunological response and for  
 CC inducing blood cell proliferation. The peptides are useful in the  
 CC treatment of disorders of the central nervous system, neurological  
 CC disorders, mental disorders, dementia, neurodegenerative diseases,  
 CC Alzheimer's disease, motor neurone disease, psychosis, neurosis, chronic  
 CC disorders of the immune system, bacterial and viral infections and  
 CC acquired immunological deficiencies  
 XX  
 SQ Sequence 15 AA;

Query Match 100.0%; Score 89; DB 4; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 9.2e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MHQPPQPLPPTVMFP 15  
 |||||  
 Db 1 MHQPPQPLPPTVMFP 15

#### RESULT 4

AAB72563

ID AAB72563 standard; peptide; 15 AA.

XX

AC AAB72563;

XX

DT 09-MAY-2001 (first entry)

XX

DE Colostrinin peptide #32.

XX

KW Neuroprotective; neural cell differentiation regulator; colostrinin;  
 KW colostrum.

XX

OS Unidentified.

XX

PN WO200112651-A2.

XX  
 PD 22-FEB-2001.  
 XX  
 PF 17-AUG-2000; 2000WO-US022774.  
 XX  
 PR 17-AUG-1999; 99US-0149633P.  
 XX  
 PA (TEXA ) UNIV TEXAS SYSTEM.  
 XX  
 PI Boldogh I;  
 XX  
 DR WPI; 2001-226545/23.  
 XX  
 PT Use of colostrinin, its constituent peptide or analog as a neural cell  
 PT regulator, for promoting neural cell differentiation and treating damaged  
 PT neural cells in a patient.  
 XX  
 PS Claim 6; Page 22; 35pp; English.  
 XX  
 CC The present invention relates to a method for promoting neural cell  
 CC differentiation and treating damaged neural cells, using colostrinin and  
 CC colostrinin constituent peptides (e.g. the present peptide) as a neural  
 CC cell regulator. Colostrinin is a polypeptide complex found in colostrum  
 XX  
 SQ Sequence 15 AA;

Query Match 100.0%; Score 89; DB 4; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 9.2e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MHQPPQPLPPTVMFP 15  
 |||||  
 Db 1 MHQPPQPLPPTVMFP 15

RESULT 5

AA014610

ID AA014610 standard; peptide; 15 AA.

XX

AC AA014610;

XX

DT 27-MAY-2002 (first entry)

XX

DE Neural cell regulatory colostrinin peptide 32.

XX

KW Neural cell differentiation; neural cell regulator; colostrinin peptide;  
 KW neural cell formation; proline-rich polypeptide aggregate; colostrum;  
 KW neural cell treatment.

XX

OS Unidentified.

XX

FH Key Location/Qualifiers

FT Modified-site 15

FT /note= "Optional C-terminal amide"

XX

PN WO200213851-A1.

XX

PD 21-FEB-2002.  
 XX  
 PF 17-AUG-2000; 2000WO-US022777.  
 XX  
 PR 17-AUG-2000; 2000WO-US022777.  
 XX  
 PA (TEXA ) UNIV TEXAS SYSTEM.  
 XX  
 PI Boldogh I, Stanton JG, Hughes TK;  
 XX  
 DR WPI; 2002-269152/31.  
 XX  
 PT Promoting cell differentiation in a patient involves use of blood cell  
 PT regulator selected from colostrinin, its constituent peptide and/or  
 PT analog.  
 XX  
 PS Claim 7; Page 22; 37pp; English.  
 XX  
 CC The invention comprises a method for promoting cell differentiation (e.g.  
 CC neural cell differentiation). The method involves contacting cells with a  
 CC neural cell regulator (i.e. a colostrinin peptide) in order to change the  
 CC cells in morphology to form neural cells. Colostrinin is a proline-rich  
 CC polypeptide aggregate that is present in colostrum. The method of the  
 CC invention is useful for promoting the differentiation of cells and for  
 CC treating damaged neural cells in a patient. The present amino acid  
 CC sequence represents a specifically claimed colostrinin peptide used in  
 CC the method of the invention  
 XX  
 SQ Sequence 15 AA;

Query Match 100.0%; Score 89; DB 5; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 9.2e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MHQPPQPLPPTVMFP 15  
 |||||  
 Db 1 MHQPPQPLPPTVMFP 15

# RESULT 6

AAM51066

ID AAM51066 standard; peptide; 15 AA.

XX

AC AAM51066;

XX

DT 30-MAY-2002 (first entry)

XX

DE Colostrinin constituent peptide (casein amino acids 159-173).

XX

KW Colostrinin; colostrum; immunomodulator; cardiovascular;  
 KW blood cell regulator; cytokine inducer; beta-casein; human.

XX

OS Homo sapiens.

XX

FH Key Location/Qualifiers

FT Modified-site 15

FT /note= "optional C-terminal amidation"

XX  
 PN WO200213849-A1.  
 XX  
 PD 21-FEB-2002.  
 XX  
 PF 17-AUG-2000; 2000WO-US022775.  
 XX  
 PR 17-AUG-2000; 2000WO-US022775.  
 XX  
 PA (TEXA ) UNIV TEXAS SYSTEM.  
 PA (REGE-) REGEN THERAPEUTICS PLC.  
 XX  
 PI Stanton GJ, Hughes TK, Boldogh I, Georgiades J;  
 XX  
 DR WPI; 2002-269150/31.  
 XX  
 PT Modulation of blood cell proliferation in a patient involves use of blood  
 PT cell regulator selected from colostrinin, its constituent peptide and/or  
 PT analog.  
 XX  
 PS Claim 1; Page 34; 54pp; English.  
 XX  
 CC The present sequence is that of a colostrinin constituent peptide that is  
 CC used as an immunological regulator and as a blood cell regulator in  
 CC claimed methods of the invention. It is classified as having a beta-  
 CC casein homologue precursor, and corresponds to casein amino acids 159-  
 CC 173. Methods are claimed for: inducing a cytokine in a cell by contact  
 CC with an immunological regulator, where the cell is present in a cell  
 CC culture, a tissue, an organ or an organism, and the cell is mammalian,  
 CC including human; modulating an immune response in a cell by contact with  
 CC the immunological regulator under conditions effective to induce a  
 CC cytokine; modulating an immune response in a patient by administering an  
 CC immunological regulator under conditions effective to induce a cytokine,  
 CC where the immunological regulator is administered topically or as part of  
 CC a dietary supplement, and where the immune response is specific or non  
 CC specific, an interferon response or an antibody response; modulating  
 CC blood cell proliferation by contacting blood cells with a blood cell  
 CC regulator, where the blood cells are present in a cell culture or an  
 CC organism, are mammalian or human, and where the blood cells are increased  
 CC in number or differentiated; and a method for modulating blood cell  
 CC proliferation in a patent. A claimed cytokine-inducing composition  
 CC comprises a pharmaceutical carrier and an active agent such as the  
 CC present peptide  
 XX  
 SQ Sequence 15 AA;

Query Match 100.0%; Score 89; DB 5; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 9.2e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHQPPQPLPPTVMFP 15  
 |||||  
 Db 1 MHQPPQPLPPTVMFP 15

RESULT 7  
 AAE20261



ID AAE20261 standard; peptide; 15 AA.  
XX  
AC AAE20261;  
XX  
DT 18-JUN-2002 (first entry)  
XX  
DE Colostrinin constituent peptide #32.  
XX  
KW Blood cell regulator; colostrinin; constituent peptide; oxidative stress;  
KW therapy; oxidative damage; skin; aging; wound healing; cell replacement;  
KW tissue; organ; cosmetic procedure; repair; regeneration; preservation;  
KW transplantation; implantation; dermatological; vulnerary.  
XX  
OS Unidentified.  
XX  
FH Key Location/Qualifiers  
FT Modified-site 15  
FT /note= "Optionally C-terminal amide"  
XX  
PN WO200213850-A1.  
XX  
PD 21-FEB-2002.  
XX  
PF 17-AUG-2000; 2000WO-US022776.  
XX  
PR 17-AUG-2000; 2000WO-US022776.  
XX  
PA (TEXA ) UNIV TEXAS SYSTEM.  
XX  
PI Stanton GJ, Hughes TK, Boldogh I;  
XX  
DR WPI; 2002-269151/31.  
XX  
PT Composition useful for the modulation of blood cell proliferation in a  
PT patient comprises a blood cell regulator selected from colostrinin, its  
PT constituent peptide and/or analog.  
XX  
PS Claim 6; Page 26; 51pp; English.  
XX  
CC The invention relates to a composition which comprises a blood cell  
CC regulator selected from colostrinin, its constituent peptide and/or  
CC analogue. The invention is used for modulating the oxidative stress level  
CC in a cell e.g. mammalian or human cell present in a cell culture, tissue,  
CC organ, or organism; or for treating oxidative damage to the skin of a  
CC patient e.g. animal or human; to modulate oxidative stress during/ after  
CC a premature birth or normal birth, preventing/delaying aging in a  
CC patient, enhancing wound healing, and the reduction of side effects of  
CC cosmetic procedures. The method changes the level of an oxidising species  
CC in the cell, such as decreases or prevents increase in the level of  
CC damage to a biomolecule of the patient selected from DNA, protein and/or  
CC lipid, compared to the same conditions when the oxidative stress  
CC regulator is not present. The modulation of oxidative stress results in  
CC enhanced repair, regeneration, and replacement of cells, tissues and  
CC organs (e.g. kidney, liver, pancreas, skin, and the other internal and  
CC external organs), as well as enhanced preservation of such organs for  
CC transplantation, implantation, or scientific research. The present  
CC sequence is a colostrinin constituent peptide

XX

SQ Sequence 15 AA;

Query Match 100.0%; Score 89; DB 5; Length 15;  
Best Local Similarity 100.0%; Pred. No. 9.2e-05;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHQPPQQLPPTVMFP 15  
|||||  
Db 1 MHQPPQQLPPTVMFP 15

# RESULT 8

AAW00679

ID AAW00679 standard; protein; 222 AA.

XX

AC AAW00679;

XX

DT 22-APR-1997 (first entry)

XX

DE Beta-casein.

XX

KW Beta-casein; goat; transgenic antithrombin III; factor XI; factor X;  
KW tAT3; serine protease inhibitor; thrombin; factor VII; factor IX; plasma;  
KW factor XII; mammary gland specific; hereditary AT3 deficiency; therapy;  
KW acquired AT3 deficiency; heparin affinity.

XX

OS Capra hircus.

XX

PN WO9626268-A1.

XX

PD 29-AUG-1996.

XX

PF 21-FEB-1996; 96WO-US002420.

XX

PR 21-FEB-1995; 95US-00391743.

XX

PA (GENZ ) GENZYME TRANSGENICS CORP.

XX

PI Ditullio P, Meade H, Cole ES;

XX

DR WPI; 1996-402361/40.

DR N-PSDB; AAT59829, AAT59830, AAT59831, AAT59832, AAT59833, AAT59834,  
DR AAT59835.

XX

PT New transgenically produced antithrombin III - useful for treating  
PT acquired or inherited AT3 deficiency, with faster clearance rate than  
PT plasma AT3.

XX

PS Disclosure; Fig 10; 37pp; English.

XX

CC This sequence represents the goat beta-casein. The gene encoding this  
CC protein was used to produce the transgenic antithrombin III (tAT3) of the  
CC invention. AT3 is a serine protease inhibitor, which inhibits thrombin  
CC and the activated forms of factors X, VII, IX, XI, and XII. The  
CC transgenic tAT3 of the invention includes a monosaccharide composition  
CC containing N-acetylgalactosamine (GalNAc). To produce a mammary gland

CC specific transgene, human AT3 cDNA (cloned as an 18.5 kb fragment) was  
 CC inserted into the goat beta-casein gene. The final 14.95 kb vector was  
 CC microinjected into goat embryos, which when developed produced the tAT3  
 CC in their milk. The tAT3 lacks O-linked glycosylation, and the major  
 CC glycoform has a complex oligosaccharide at each glycosylation site,  
 CC except the Asn155 residue. At Asn155, it has a significant amount of  
 CC oligomannose, and hybrid residues. The tAT3 is used in the same way as  
 CC plasma AT3, for the treatment of hereditary or acquired AT3 deficiency.  
 CC The tAT3, however, has a faster clearance rate than plasma AT3 and may  
 CC also have a higher affinity for heparin. The tAT3 can also be produced  
 CC without the variability, immunogenicity and viral contamination problems  
 CC associated with plasma-derived material

XX

SQ Sequence 222 AA;

Query Match 91.0%; Score 81; DB 2; Length 222;

Best Local Similarity 93.3%; Pred. No. 0.01;

Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MHQPPQPLPPTVMFP 15

||||||| |||||

Db 159 MHQPPQPLSPTVMFP 173

RESULT 9

AAR80281

ID AAR80281 standard; protein; 209 AA.

XX

AC AAR80281;

XX

DT 14-FEB-1996 (first entry)

XX

DE Methyl or ethyl esterified bovine beta-casein A1.

XX

KW Bovine; beta-casein; ethyl esterification; pepsin hydrolysis;

KW proteolysis; peptide ester; food; pharmaceutical; cosmetics.

XX

OS Bos taurus.

XX

FH Key Location/Qualifiers

FT Protein 1. .209

FT /note= "55% esterified by methanol or by ethanol,

FT resulting in atypical pepsin cleavage sites, in addition

FT to the naturally occurring (native) sites"

FT Peptide 2. .25

FT /label= A

FT /note= "tryptic peptide from native protein"

FT Cleavage-site 4. .5

FT /note= "pepsin cleavage site in native protein"

FT Cleavage-site 5. .6

FT /note= "pepsin cleavage site in native protein and in

FT methyl ester of beta-casein"

FT Cleavage-site 11. .12

FT /note= "newly identified pepsin cleavage site in methyl

FT ester of beta-casein"

FT Cleavage-site 15. .16

FT /note= "pepsin cleavage site in native protein"

FT	Modified-site	15
FT		/note= "phosphorylated"
FT	Modified-site	17
FT		/note= "phosphorylated"
FT	Modified-site	18
FT		/note= "phosphorylated"
FT	Modified-site	19
FT		/note= "phosphorylated"
FT	Peptide	26. .28
FT		/label= B
FT		/note= "tryptic peptide from native protein"
FT	Peptide	29. .32
FT		/label= C
FT		/note= "tryptic peptide from native protein"
FT	Peptide	33. .48
FT		/label= D
FT		/note= "tryptic peptide from native protein"
FT	Modified-site	35
FT		/note= "phosphorylated"
FT	Cleavage-site	44. .45
FT		/note= "pepsin cleavage site in native protein"
FT	Cleavage-site	45. .46
FT		/note= "pepsin cleavage site in native protein and in ethyl ester of beta-casein"
FT	Peptide	49. .97
FT		/label= E
FT		/note= "tryptic peptide from native protein"
FT	Cleavage-site	55. .56
FT		/note= "pepsin cleavage site in native protein and in ethyl ester of beta-casein"
FT	Cleavage-site	57. .58
FT		/note= "pepsin cleavage site in native protein and in ethyl ester of beta-casein"
FT	Cleavage-site	58. .59
FT		/note= "pepsin cleavage site in native protein and in ethyl ester of beta-casein"
FT	Cleavage-site	72. .73
FT		/note= "pepsin cleavage site in native protein"
FT	Cleavage-site	73. .74
FT		/note= "newly identified pepsin cleavage site in methyl ester of beta-casein"
FT	Cleavage-site	80. .81
FT		/note= "pepsin cleavage site in native protein and in ethyl ester of beta-casein"
FT	Cleavage-site	93. .94
FT		/note= "pepsin cleavage site in native protein and in methyl and ethyl esters of beta-casein"
FT	Peptide	100. .105
FT		/label= F
FT		/note= "tryptic peptide from native protein"
FT	Peptide	106. .107
FT		/label= G
FT		/note= "tryptic peptide from native protein"
FT	Peptide	108. .113
FT		/label= H
FT		/note= "tryptic peptide from native protein"
FT	Peptide	114. .169

```

FT          /label= I
FT          /note= "tryptic peptide from native protein"
FT  Cleavage-site  125. .126
FT          /note= "pepsin cleavage site in native protein"
FT  Cleavage-site  126. .127
FT          /note= "pepsin cleavage site in native protein"
FT  Cleavage-site  127. .128
FT          /note= "pepsin cleavage site in native protein"
FT  Cleavage-site  141. .142
FT          /note= "pepsin cleavage site in native protein"
FT  Cleavage-site  142. .143
FT          /note= "pepsin cleavage site in native protein and in
FT          methyl and ethyl esters of beta-casein"
FT  Cleavage-site  156. .157
FT          /note= "newly identified pepsin cleavage site in ethyl
FT          ester of beta-casein"
FT  Cleavage-site  162. .163
FT          /note= "newly identified pepsin cleavage site in ethyl
FT          ester of beta-casein"
FT  Cleavage-site  163. .164
FT          /note= "pepsin cleavage site in native protein and in
FT          methyl and ethyl esters of beta-casein"
FT  Cleavage-site  164. .165
FT          /note= "pepsin cleavage site in native protein and in
FT          ethyl ester of beta-casein"
FT  Peptide        170. .176
FT          /label= J
FT          /note= "tryptic peptide from native protein"
FT  Peptide        177. .183
FT          /label= K
FT          /note= "tryptic peptide from native protein"
FT  Peptide        184. .202
FT          /label= L
FT          /note= "tryptic peptide from native protein"
FT  Cleavage-site  188. .189
FT          /note= "pepsin cleavage site in native protein and in
FT          methyl ester of beta-casein"
FT  Cleavage-site  189. .190
FT          /note= "pepsin cleavage site in native protein and in
FT          methyl and ethyl esters of beta-casein"
FT  Cleavage-site  190. .191
FT          /note= "pepsin cleavage site in native protein and in
FT          ethyl ester of beta-casein"
FT  Cleavage-site  191. .192
FT          /note= "pepsin cleavage site in native protein and in
FT          methyl and ethyl esters of beta-casein"
FT  Cleavage-site  192. .193
FT          /note= "pepsin cleavage site in native protein and in
FT          methyl and ethyl esters of beta-casein"
FT  Cleavage-site  198. .199
FT          /note= "newly identified pepsin cleavage site in methyl
FT          and ethyl esters of beta-casein"
FT  Peptide        203. .209
FT          /label= N
FT          /note= "tryptic peptide from native protein"
FT  Cleavage-site  207. .208
FT          /note= "newly identified pepsin cleavage site in methyl

```

FT ester of beta-casein"

XX

PN WO9517518-A1.

XX

PD 29-JUN-1995.

XX

PF 20-DEC-1994; 94WO-FR001500.

XX

PR 23-DEC-1993; 93FR-00015764.

XX

PA (INRG ) INST NAT RECH AGRONOMIQUE.

XX

PI Chobert J, Briand L, Haertle T;

XX

DR WPI; 1995-240679/31.

XX

PT New esterified amino acids, peptide(s) and their mixts. - prepd. by  
PT esterification of protein then enzymatic hydrolysis, useful as  
PT ingredients and additives in foods, pharmaceuticals and cosmetics.

XX

PS Claim 7; Fig 7 and 18; 47pp; French.

XX

CC The native form of bovine beta-casein A1 contains various pepsin cleavage  
CC sites. Esterification of the protein with methanol or ethanol results in  
CC a form of beta-casein contg. additional, non-conventional pepsin cleavage  
CC sites (see Features Table). Esterified peptides and amino acids (and  
CC their mixtures) resulting from hydrolysis of an esterified protein (pref.  
CC beta-lactoglobulin or beta-casein) are claimed. The hydrolysis products  
CC are useful as ingredients, additives or active agents in foods,  
CC pharmaceuticals and cosmetics

XX

SQ Sequence 209 AA;

Query Match 89.9%; Score 80; DB 2; Length 209;

Best Local Similarity 93.3%; Pred. No. 0.013;

Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MHQPPQPLPPTVMFP 15

|||| |||

Db 144 MHQPHQPLPPTVMFP 158

RESULT 10

AAR95609

ID AAR95609 standard; protein; 209 AA.

XX

AC AAR95609;

XX

DT 26-NOV-1996 (first entry)

XX

DE Bovine beta casein A1 variant.

XX

KW Milk; beta casein; diabetogenic; diabetes; cow; milk products; butter;  
KW cheese; cream.

XX

OS Bos taurus.

XX

FH Key Location/Qualifiers  
 FT Region 63. .68  
 FT /label= Diabetogenic hexapeptide.  
 XX  
 PN WO9614577-A1.  
 XX  
 PD 17-MAY-1996.  
 XX  
 PF 03-NOV-1995; 95WO-NZ000114.  
 XX  
 PR 04-NOV-1994; 94NZ-00264862.  
 XX  
 PA (NACH-) NAT CHILD HEALTH RES FOUND.  
 PA (NZDA-) NEW ZEALAND DAIRY BOARD.  
 XX  
 PI Elliott RB, Hill JP;  
 XX  
 DR WPI; 1996-251885/25.  
 XX  
 PT Selecting non-diabetogenic milk and milk prods. - by testing milk or cows  
 PT for the presence of non-diabetogenic variants of beta-casein.  
 XX  
 PS Disclosure; Fig 2; 28pp; English.  
 XX  
 CC A method for selecting milk for feeding to diabetes susceptible individuals  
 CC comprises testing milk from identified cows for the presence of variants  
 CC of beta casein and selecting those cows whose milk contains non-  
 CC diabetogenic variants and milking these cows separately. The milk and  
 CC milk products obtained can reduce the risk of susceptible individuals  
 CC contracting Type-1 diabetes  
 XX  
 SQ Sequence 209 AA;

Query Match 89.9%; Score 80; DB 2; Length 209;  
 Best Local Similarity 93.3%; Pred. No. 0.013;  
 Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MHQPPQPLPPTVMFP 15  
 |||| |||||  
 Db 144 MHQPHQPLPPTVMFP 158

# RESULT 11

AAW54021

ID AAW54021 standard; protein; 209 AA.

XX

AC AAW54021;

XX

DT 31-JUL-1998 (first entry)

XX

DE Modified Casein sequence.

XX

KW Casein; cow; mineral solubilising; A2 beta-casein variant.

XX

OS Bos sp.

XX

FH Key Location/Qualifiers

FT Modified-site 15  
 FT /note= "phosphorylated"  
 FT Modified-site 17  
 FT /note= "phosphorylated"  
 FT Modified-site 18  
 FT /note= "phosphorylated"  
 FT Modified-site 19  
 FT /note= "phosphorylated"  
 FT Misc-difference 25  
 FT /label= R25C  
 FT Misc-difference 88  
 FT /label= L88I  
 FT Misc-difference 117  
 FT /label= Q117E  
 FT Misc-difference 175  
 FT /label= E175Q  
 FT Misc-difference 195  
 FT /label= Q195E

XX

PN WO9812222-A1.

XX

PD 26-MAR-1998.

XX

PF 23-SEP-1997; 97WO-KR000182.

XX

PR 23-SEP-1996; 96KR-00043482.

XX

PA (HANS/) HAN S K.

XX

PI Shin YC;

XX

DR WPI; 1998-217207/19.

XX

PT New casein and its preparation - has a higher mineral solubilising effect.

XX

PS Claim 1; Page 23-24; 32pp; English.

XX

CC This sequence represents the modified casein protein of the invention. It  
 CC is based on the bovine casein sequence and has the following  
 CC substitutions in comparison with the conventional A2 variant of beta-  
 CC casein: R25C, L88I, Q117E, E175Q and Q195E. The new casein has a higher  
 CC mineral solubilising effect and is therefore more effective than previous  
 CC caseins

XX

SQ Sequence 209 AA;

Query Match 89.9%; Score 80; DB 2; Length 209;

Best Local Similarity 93.3%; Pred. No. 0.013;

Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MHQPPQPLPPTVMFP 15  
 |||| |||||  
 Db 144 MHQPHQPLPPTVMFP 158

RESULT 12



AAB60348

ID AAB60348 standard; protein; 209 AA.

XX

AC AAB60348;

XX

DT 06-APR-2001 (first entry)

XX

DE Bovine beta-casein type A2.

XX

KW Bovine; beta-casein; type A2; milk; dairy product; dietary supplement;  
KW plasma homocysteine; homocystine; tHcy; folic acid; pyridoxine; betaine;  
KW cobalamin; vascular disease; diabetes; peripheral vascular disease;  
KW blood vessel wall degeneration; cardiovascular disease;  
KW cerebrovascular disease; neural tube defect; prophylaxis.

XX

OS Bos taurus.

XX

PN WO200100047-A1.

XX

PD 04-JAN-2001.

XX

PF 29-JUN-2000; 2000WO-NZ000116.

XX

PR 29-JUN-1999; 99NZ-00336505.

PR 18-APR-2000; 2000NZ-00504057.

XX

PA (NZMI-) NEW ZEALAND MILK INST LTD.

XX

PI Elliott RB, Laugesen BM;

XX

DR WPI; 2001-112363/12.

XX

PT Prophylactic dietary supplement for reducing incidence of cardio, cerebro  
PT vascular diseases and diabetes in a population, contains milk or its  
PT products fortified with betaine, cobalamin, folic acid or pyridoxine.

XX

PS Claim 9; Fig 3; 32pp; English.

XX

CC The invention relates to a dietary supplement which, when consumed, is  
CC capable of reducing plasma levels of homocyst(e)ine (tHcy). tHcy is a  
CC major risk indicator of heart disease and vascular disease in general in  
CC humans. Vascular wall health is also seriously compromised in patients  
CC with clinical or unrecognised diabetes, with tHcy being a strong risk  
CC factor for mortality in type II diabetic patients. Deficiencies in folic  
CC acid, pyridoxine and cobalamin lead to higher tHcy levels, and folic acid  
CC deficiency is known to be involved in vascular disease, as well as  
CC causing neural tube defects in early embryonic development. The dietary  
CC supplement of the invention comprises milk or a milk product, fortified  
CC by the addition of betaine, cobalamin, folic acid, pyridoxine or their  
CC analogues. In addition, the beta-casein component of the milk is  
CC substantially the A2 variant. Beta-casein types A1 and B, consumption of  
CC which are correlated with the incidence of type I diabetes, are  
CC substantially excluded from the supplement. The dietary supplement is  
CC useful for reducing the incidence of vascular disease, including  
CC peripheral vascular disease and blood vessel wall degeneration and  
CC particularly cardiovascular disease and cerebrovascular disease, and is  
CC also useful for reducing the incidence of type I and II diabetes. It

CC additionally provides a sufficient daily dose of folic acid to prevent  
CC neural tube defects in fetuses. The supplement provides health  
CC improvements to a human population without the administration of  
CC medication. The present sequence represents bovine beta-casein type A2  
XX  
SQ Sequence 209 AA;

Query Match 83.1%; Score 74; DB 4; Length 209;  
Best Local Similarity 80.0%; Pred. No. 0.075;  
Matches 12; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MHQPPQPLPPTVMFP 15  
||:| :|||||||  
Db 144 MHEPHEPLPPTVMFP 158

RESULT 13

AAB72523

ID AAB72523 standard; peptide; 10 AA.

XX

AC AAB72523;

XX

DT 09-MAY-2001 (first entry)

XX

DE Colostrinin peptide #24.

XX

KW Dermatological; oxidative stress regulator; colostrinin.

XX

OS Unidentified.

XX

PN WO200112650-A2.

XX

PD 22-FEB-2001.

XX

PF 17-AUG-2000; 2000WO-US022665.

XX

PR 17-AUG-1999; 99US-0149310P.

XX

PA (TEXA ) UNIV TEXAS SYSTEM.

XX

PI Stanton GJ, Hughes TK, Boldogh I;

XX

DR WPI; 2001-218342/22.

XX

PT Modulating oxidative stress level in a cell, involves contacting the cell  
PT with an oxidative stress regulator selected from colostrinin, its  
PT constituent peptide, analog or their combinations.

XX

PS Claim 6; Page 26; 48pp; English.

XX

CC The present invention relates to a method for modulating the oxidative  
CC stress level in a cell or a patient, comprising contacting the cell with,  
CC or administering to the patient, an oxidative stress regulator selected  
CC from colostrinin, or its constituent peptide (e.g. the present peptide),  
CC to change the level of an oxidising species in the cell. The method can  
CC be used to treat oxidative damage to skin, by decreasing or preventing an  
CC increase in the level of damage to a biomolecule of the patient

XX  
SQ Sequence 10 AA;

Query Match 64.0%; Score 57; DB 4; Length 10;  
Best Local Similarity 100.0%; Pred. No. 0.66;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 QPLPPTVMFP 15  
| | | | | | | |  
Db 1 QPLPPTVMFP 10

RESULT 14

AAB72270

ID AAB72270 standard; peptide; 10 AA.

XX

AC AAB72270;

XX

DT 14-MAY-2001 (first entry)

XX

DE Colostrinin derived cytokine inducing peptide SEQ ID 25.

XX

KW Colostrinin; immune response; cytokine; blood cell proliferation;  
KW central nervous system disorder; neurological disorder; mental disorder;  
KW dementia; neurodegenerative disease; Alzheimer's disease; psychosis;  
KW neurosis; infection.

XX

OS Synthetic.

XX

PN WO200111937-A2.

XX

PD 22-FEB-2001.

XX

PF 17-AUG-2000; 2000WO-US022818.

XX

PR 17-AUG-1999; 99US-0149311P.

XX

PA (TEXA ) UNIV TEXAS SYSTEM.

PA (REGE-) REGEN THERAPEUTICS PLC.

XX

PI Stanton GJ, Hughes TK, Boldogh I, Georgiades J;

XX

DR WPI; 2001-202804/20.

XX

PT Inducing a cytokine and modulating an immune response, useful for  
PT treating central nervous system diseases and bacterial and viral  
PT infections, comprises administering colostrinin as an immunological  
PT regulator.

XX

PS Claim 1; Page 34; 50pp; English.

XX

CC Sequences AAB72246 - AAB72275 represent peptides derived from colostrinin,  
CC a proline rich polypeptide aggregate contained in colostrum. The peptides  
CC have immune response modulatory activity, and are capable of inducing  
CC cytokines. Colostrinin and its derived peptides are useful for inducing  
CC cytokine production, for modulating an immunological response and for  
CC inducing blood cell proliferation. The peptides are useful in the

CC treatment of disorders of the central nervous system, neurological  
CC disorders, mental disorders, dementia, neurodegenerative diseases,  
CC Alzheimer's disease, motor neurone disease, psychosis, neurosis, chronic  
CC disorders of the immune system, bacterial and viral infections and  
CC acquired immunological deficiencies

XX

SQ Sequence 10 AA;

Query Match 64.0%; Score 57; DB 4; Length 10;  
Best Local Similarity 100.0%; Pred. No. 0.66;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 QPLPPTVMFP 15

|||||||

Db 1 QPLPPTVMFP 10

RESULT 15

AAB72555

ID AAB72555 standard; peptide; 10 AA.

XX

AC AAB72555;

XX

DT 09-MAY-2001 (first entry)

XX

DE Colostrinin peptide #24.

XX

KW Neuroprotective; neural cell differentiation regulator; colostrinin;

KW colostrum.

XX

OS Unidentified.

XX

PN WO200112651-A2.

XX

PD 22-FEB-2001.

XX

PF 17-AUG-2000; 2000WO-US022774.

XX

PR 17-AUG-1999; 99US-0149633P.

XX

PA (TEXA ) UNIV TEXAS SYSTEM.

XX

PI Boldogh I;

XX

DR WPI; 2001-226545/23.

XX

PT Use of colostrinin, its constituent peptide or analog as a neural cell

PT regulator, for promoting neural cell differentiation and treating damaged

PT neural cells in a patient.

XX

PS Claim 6; Page 21; 35pp; English.

XX

CC The present invention relates to a method for promoting neural cell

CC differentiation and treating damaged neural cells, using colostrinin and

CC colostrinin constituent peptides (e.g. the present peptide) as a neural

CC cell regulator. Colostrinin is a polypeptide complex found in colostrum

XX

SQ Sequence 10 AA;

Query Match 64.0%; Score 57; DB 4; Length 10;  
Best Local Similarity 100.0%; Pred. No. 0.66;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 QPLPPTVMFP 15

|||||||

Db 1 QPLPPTVMFP 10

Search completed: August 24, 2004, 15:42:49  
Job time : 63.1194 secs

OM protein - protein search, using sw model

Run on: August 24, 2004, 15:33:13 ; Search time 16.4552 Seconds  
 (without alignments)  
 47.060 Million cell updates/sec

Title: US-09-641-801-34  
 Perfect score: 89  
 Sequence: 1 MHQPPQPLPPTVMFP 15

Scoring table: BLOSUM62  
 Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
 Maximum Match 100%  
 Listing first 45 summaries

Database : Issued Patents\_AA:\*  
 1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep:\*  
 2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep:\*  
 3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep:\*  
 4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep:\*  
 5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep:\*  
 6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
 score greater than or equal to the score of the result being printed,  
 and is derived by analysis of the total score distribution.

# SUMMARIES

		%					
Result		Query					
No.	Score	Match	Length	DB	ID	Description	
1	89	100.0	15	4	US-09-641-803-34	Sequence 34, Appl	
2	81	91.0	222	2	US-08-391-743A-2	Sequence 2, Appli	
3	81	91.0	222	4	US-09-143-155-2	Sequence 2, Appli	
4	80	89.9	209	3	US-09-269-220-1	Sequence 1, Appli	
5	80	89.9	209	3	US-09-269-220-2	Sequence 2, Appli	
6	80	89.9	209	4	US-08-836-778-2	Sequence 2, Appli	
7	57	64.0	10	4	US-09-641-803-25	Sequence 25, Appl	
8	53	59.6	10	4	US-09-794-346-1	Sequence 1, Appli	
9	50	56.2	751	4	US-10-020-079-8	Sequence 8, Appli	
10	50	56.2	764	4	US-10-020-079-6	Sequence 6, Appli	
11	50	56.2	776	4	US-10-020-079-24	Sequence 24, Appl	

12	50	56.2	789	4	US-10-020-079-22	Sequence 22, Appl
13	50	56.2	838	4	US-10-020-079-40	Sequence 40, Appl
14	50	56.2	851	4	US-10-020-079-38	Sequence 38, Appl
15	50	56.2	863	4	US-10-020-079-32	Sequence 32, Appl
16	50	56.2	864	4	US-10-020-079-4	Sequence 4, Appli
17	50	56.2	870	4	US-10-020-079-2	Sequence 2, Appli
18	50	56.2	876	4	US-10-020-079-30	Sequence 30, Appl
19	50	56.2	889	4	US-10-020-079-20	Sequence 20, Appl
20	50	56.2	895	4	US-10-020-079-18	Sequence 18, Appl
21	50	56.2	951	4	US-10-020-079-36	Sequence 36, Appl
22	50	56.2	957	4	US-10-020-079-34	Sequence 34, Appl
23	50	56.2	976	4	US-10-020-079-28	Sequence 28, Appl
24	50	56.2	982	4	US-10-020-079-26	Sequence 26, Appl
25	47	52.8	188	4	US-09-252-991A-28564	Sequence 28564, A
26	47	52.8	210	1	US-08-078-090-2	Sequence 2, Appli
27	47	52.8	213	3	US-09-131-028A-2	Sequence 2, Appli
28	47	52.8	213	3	US-09-131-028A-12	Sequence 12, Appl
29	47	52.8	385	1	US-08-450-257-58	Sequence 58, Appl
30	47	52.8	385	1	US-08-450-246-58	Sequence 58, Appl
31	47	52.8	385	1	US-08-450-098-58	Sequence 58, Appl
32	47	52.8	385	1	US-08-451-233-58	Sequence 58, Appl
33	47	52.8	385	1	US-08-450-236-58	Sequence 58, Appl
34	47	52.8	385	4	US-08-235-403-58	Sequence 58, Appl
35	47	52.8	490	3	US-09-039-555B-14	Sequence 14, Appl
36	46	51.7	154	4	US-09-252-991A-30960	Sequence 30960, A
37	46	51.7	480	4	US-09-149-476-405	Sequence 405, App
38	46	51.7	708	4	US-09-857-556A-12	Sequence 12, Appl
39	45	50.6	97	4	US-09-673-395A-388	Sequence 388, App
40	45	50.6	173	4	US-09-252-991A-23800	Sequence 23800, A
41	45	50.6	278	4	US-09-252-991A-32988	Sequence 32988, A
42	45	50.6	2414	1	US-08-227-536-2	Sequence 2, Appli
43	45	50.6	2414	5	PCT-US95-04682-2	Sequence 2, Appli
44	44	49.4	95	4	US-09-314-268-132	Sequence 132, App
45	43.5	48.9	255	4	US-09-489-039A-9101	Sequence 9101, Ap

#### ALIGNMENTS

##### RESULT 1

US-09-641-803-34

; Sequence 34, Application US/09641803

; Patent No. 6500798

##### ; GENERAL INFORMATION:

; APPLICANT: STANTON, G. John

; APPLICANT: HUGHES, Thomas K.

; APPLICANT: BOLDOGH, Istvan

; TITLE OF INVENTION: USE OF COLOSTRININ, CONSTITUENT PEPTIDES THEREOF, AND

; TITLE OF INVENTION: ANALOGS THEREOF AS OXIDATIVE STRESS REGULATORS

; FILE REFERENCE: 265.00220101

; CURRENT APPLICATION NUMBER: US/09/641,803

; CURRENT FILING DATE: 2000-08-17

; PRIOR APPLICATION NUMBER: 60/149,310

; PRIOR FILING DATE: 1999-08-17

; NUMBER OF SEQ ID NOS: 34

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 34

; LENGTH: 15  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: synthetic  
; OTHER INFORMATION: peptide  
US-09-641-803-34

Query Match 100.0%; Score 89; DB 4; Length 15;  
Best Local Similarity 100.0%; Pred. No. 3.2e-05;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MHQPPQPLPPTVMFP 15  
| | | | | | | | | | | | | | | |  
Db 1 MHQPPQPLPPTVMFP 15

RESULT 2

US-08-391-743A-2

; Sequence 2, Application US/08391743A  
; Patent No. 5843705  
; GENERAL INFORMATION:  
; APPLICANT: DiTullio, Paul A.; Meade, Harry; Cole, Edward S.  
; TITLE OF INVENTION: TRANSGENETICALLY PRODUCED ANTITHROMBIN III  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: LAHIVE & COCKFIELD, LLP  
; STREET: 28 State Street  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/391,743A  
; FILING DATE: 21-FEB-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Myers, Paul Louis  
; REGISTRATION NUMBER: 35,965  
; REFERENCE/DOCKET NUMBER: TCI-045  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617)227-7400  
; TELEFAX: (617)742-4214  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 222 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein



US-08-391-743A-2

Query Match 91.0%; Score 81; DB 2; Length 222;  
Best Local Similarity 93.3%; Pred. No. 0.0036;  
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MHQPPQPLPPTVMFP 15  
||||||| |||||  
Db 159 MHQPPQPLSPTVMFP 173

RESULT 3

US-09-143-155-2

; Sequence 2, Application US/09143155  
; Patent No. 6441145  
; GENERAL INFORMATION:  
; APPLICANT: DiTullio, Paul A.; Meade, Harry; Cole, Edward S.  
; TITLE OF INVENTION: TRANSGENETICALLY PRODUCED ANTITHROMBIN III  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: LAHIVE & COCKFIELD, LLP  
; STREET: 28 State Street  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/143,155  
; FILING DATE: 28-Aug-1998  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/391,743  
; FILING DATE: <Unknown>  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Myers, Paul Louis  
; REGISTRATION NUMBER: 35,965  
; REFERENCE/DOCKET NUMBER: TCI-045  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617)227-7400  
; TELEFAX: (617)742-4214  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 222 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:

US-09-143-155-2

Query Match 91.0%; Score 81; DB 4; Length 222;  
Best Local Similarity 93.3%; Pred. No. 0.0036;  
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MHQPPQPLPPTVMFP 15  
 ||||| |||||  
 Db 159 MHQPPQPLSPTVMFP 173

RESULT 4

US-09-269-220-1

; Sequence 1, Application US/09269220  
 ; Patent No. 6180761  
 ; GENERAL INFORMATION:  
 ; APPLICANT: HAN, Sang K  
 ; APPLICANT: SHIN, Yoo C  
 ; TITLE OF INVENTION: CASEIN AND PROCESS FOR THE PREPARATION THEREOF  
 ; FILE REFERENCE: 1423.1001/MJH  
 ; CURRENT APPLICATION NUMBER: US/09/269,220  
 ; CURRENT FILING DATE: 1999-03-23  
 ; PRIOR APPLICATION NUMBER: KR 1996-43482  
 ; PRIOR FILING DATE: 1996-03-23  
 ; PRIOR APPLICATION NUMBER: PCT/KR97/00182  
 ; PRIOR FILING DATE: 1997-09-23  
 ; NUMBER OF SEQ ID NOS: 7  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 1  
 ; LENGTH: 209  
 ; TYPE: PRT  
 ; ORGANISM: Bos taurus  
 ; FEATURE:  
 ; NAME/KEY: ACT\_SITE  
 ; LOCATION: (15)  
 ; OTHER INFORMATION: phospholyated serine  
 ; NAME/KEY: ACT\_SITE  
 ; LOCATION: (17)..(19)  
 US-09-269-220-1

Query Match 89.9%; Score 80; DB 3; Length 209;  
 Best Local Similarity 93.3%; Pred. No. 0.0046;  
 Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MHQPPQPLPPTVMFP 15  
 |||| |||||  
 Db 144 MHQPHQPLPPTVMFP 158

RESULT 5

US-09-269-220-2

; Sequence 2, Application US/09269220  
 ; Patent No. 6180761  
 ; GENERAL INFORMATION:  
 ; APPLICANT: HAN, Sang K  
 ; APPLICANT: SHIN, Yoo C  
 ; TITLE OF INVENTION: CASEIN AND PROCESS FOR THE PREPARATION THEREOF  
 ; FILE REFERENCE: 1423.1001/MJH  
 ; CURRENT APPLICATION NUMBER: US/09/269,220  
 ; CURRENT FILING DATE: 1999-03-23  
 ; PRIOR APPLICATION NUMBER: KR 1996-43482  
 ; PRIOR FILING DATE: 1996-03-23

; PRIOR APPLICATION NUMBER: PCT/KR97/00182  
; PRIOR FILING DATE: 1997-09-23  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 209  
; TYPE: PRT  
; ORGANISM: Bos taurus  
US-09-269-220-2

Query Match 89.9%; Score 80; DB 3; Length 209;  
Best Local Similarity 93.3%; Pred. No. 0.0046;  
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MHQPPQPLPPTVMFP 15  
|||| |||||  
Db 144 MHQPHQPLPPTVMFP 158

RESULT 6  
US-08-836-778-2  
; Sequence 2, Application US/08836778  
; Patent No. 6451368  
; GENERAL INFORMATION:  
; APPLICANT: ELLIOTT, ROBERT BARTLETT  
; APPLICANT: HILL, JEREMY PAUL  
; TITLE OF INVENTION: METHOD OF SELECTING NON-DIABETOGENIC MILK OR MILK  
; TITLE OF INVENTION: PRODUCTS AND MILK OR MILK PRODUCTS SO SELECTED  
; FILE REFERENCE: P369648 DCC  
; CURRENT APPLICATION NUMBER: US/08/836,778  
; CURRENT FILING DATE: 1995-11-03  
; PRIOR APPLICATION NUMBER: NZ 264862  
; PRIOR FILING DATE: 1994-11-04  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 209  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence:BOVINE MILK  
; OTHER INFORMATION: PROTEIN  
US-08-836-778-2

Query Match 89.9%; Score 80; DB 4; Length 209;  
Best Local Similarity 93.3%; Pred. No. 0.0046;  
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MHQPPQPLPPTVMFP 15  
|||| |||||  
Db 144 MHQPHQPLPPTVMFP 158

RESULT 7  
US-09-641-803-25  
; Sequence 25, Application US/09641803  
; Patent No. 6500798

```
; GENERAL INFORMATION:
; APPLICANT: STANTON, G. John
; APPLICANT: HUGHES, Thomas K.
; APPLICANT: BOLDOGH, Istvan
; TITLE OF INVENTION: USE OF COLOSTRININ, CONSTITUENT PEPTIDES THEREOF, AND
; TITLE OF INVENTION: ANALOGS THEREOF AS OXIDATIVE STRESS REGULATORS
; FILE REFERENCE: 265.00220101
; CURRENT APPLICATION NUMBER: US/09/641,803
; CURRENT FILING DATE: 2000-08-17
; PRIOR APPLICATION NUMBER: 60/149,310
; PRIOR FILING DATE: 1999-08-17
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 25
; LENGTH: 10
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: peptide
US-09-641-803-25
```

```
Query Match          64.0%; Score 57; DB 4; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.18;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy          6 QPLPPTVMFP 15
             |||||
Db          1 QPLPPTVMFP 10
```

```
RESULT 8
US-09-794-346-1
; Sequence 1, Application US/09794346
; Patent No. 6627604
; GENERAL INFORMATION:
; APPLICANT: Aventis Pharma Deutschland GmbH
; TITLE OF INVENTION: Memno Peptides, Process for Their Preparation and Use
; Thereof
; FILE REFERENCE: 02481.1728
; CURRENT APPLICATION NUMBER: US/09/794,346
; CURRENT FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: EP 00104114.4
; PRIOR FILING DATE: 2000-02-29
; PRIOR APPLICATION NUMBER: PCT/EP 01/01661
; PRIOR FILING DATE: 2001-02-15
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 10
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Description of Artificial Sequence: Memnoniella
echinata, FH 227
; OTHER INFORMATION: 1, DSM 1319
```

US-09-794-346-1

Query Match 59.6%; Score 53; DB 4; Length 10;  
Best Local Similarity 90.0%; Pred. No. 0.55;  
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MHQPPQPLPP 10  
      |||| |||||  
Db 1 MHQPHQPLPP 10

RESULT 9

US-10-020-079-8

; Sequence 8, Application US/10020079  
; Patent No. 6579710  
; GENERAL INFORMATION:  
; APPLICANT: Turner, C. Alexander Jr.  
; APPLICANT: Mathur, Brian  
; APPLICANT: Friddle, Carl Johan  
; TITLE OF INVENTION: No. 6579710e1 Human Kinases and Polynucleotides Encoding  
the Same  
; FILE REFERENCE: LEX-0281-USA  
; CURRENT APPLICATION NUMBER: US/10/020,079  
; CURRENT FILING DATE: 2001-12-12  
; PRIOR APPLICATION NUMBER: US 60/255,103  
; PRIOR FILING DATE: 2000-12-12  
; PRIOR APPLICATION NUMBER: US 60/289,422  
; PRIOR FILING DATE: 2001-05-08  
; NUMBER OF SEQ ID NOS: 40  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 8  
; LENGTH: 751  
; TYPE: PRT  
; ORGANISM: homo sapiens  
; FEATURE:  
; NAME/KEY: VARIANT  
; LOCATION: (1)...(751)  
; OTHER INFORMATION: Xaa = Any Amino Acid  
US-10-020-079-8

Query Match 56.2%; Score 50; DB 4; Length 751;  
Best Local Similarity 88.9%; Pred. No. 70;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 HQPPQPLPP 10  
      | |||||  
Db 597 HLPPQPLPP 605

RESULT 10

US-10-020-079-6

; Sequence 6, Application US/10020079  
; Patent No. 6579710  
; GENERAL INFORMATION:  
; APPLICANT: Turner, C. Alexander Jr.  
; APPLICANT: Mathur, Brian  
; APPLICANT: Friddle, Carl Johan

```
; TITLE OF INVENTION: No. 6579710e1 Human Kinases and Polynucleotides Encoding
the Same
; FILE REFERENCE: LEX-0281-USA
; CURRENT APPLICATION NUMBER: US/10/020,079
; CURRENT FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: US 60/255,103
; PRIOR FILING DATE: 2000-12-12
; PRIOR APPLICATION NUMBER: US 60/289,422
; PRIOR FILING DATE: 2001-05-08
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 764
; TYPE: PRT
; ORGANISM: homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(764)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-10-020-079-6
```

```
Query Match          56.2%; Score 50; DB 4; Length 764;
Best Local Similarity 88.9%; Pred. No. 71;
Matches      8; Conservative    0; Mismatches    1; Indels      0; Gaps      0;
```

```
Qy          2 HQPPQPLPP 10
              | |||||
Db          610 HLPPQPLPP 618
```

RESULT 11

```
US-10-020-079-24
; Sequence 24, Application US/10020079
; Patent No. 6579710
; GENERAL INFORMATION:
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Mathur, Brian
; APPLICANT: Friddle, Carl Johan
; TITLE OF INVENTION: No. 6579710e1 Human Kinases and Polynucleotides Encoding
the Same
; FILE REFERENCE: LEX-0281-USA
; CURRENT APPLICATION NUMBER: US/10/020,079
; CURRENT FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: US 60/255,103
; PRIOR FILING DATE: 2000-12-12
; PRIOR APPLICATION NUMBER: US 60/289,422
; PRIOR FILING DATE: 2001-05-08
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 24
; LENGTH: 776
; TYPE: PRT
; ORGANISM: homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(776)
; OTHER INFORMATION: Xaa = Any Amino Acid
```

US-10-020-079-24

Query Match 56.2%; Score 50; DB 4; Length 776;  
Best Local Similarity 88.9%; Pred. No. 72;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 HQPPQPLPP 10  
| |||||  
Db 597 HLPPQPLPP 605

RESULT 12

US-10-020-079-22

; Sequence 22, Application US/10020079  
; Patent No. 6579710  
; GENERAL INFORMATION:  
; APPLICANT: Turner, C. Alexander Jr.  
; APPLICANT: Mathur, Brian  
; APPLICANT: Friddle, Carl Johan  
; TITLE OF INVENTION: No. 6579710e1 Human Kinases and Polynucleotides Encoding  
the Same  
; FILE REFERENCE: LEX-0281-USA  
; CURRENT APPLICATION NUMBER: US/10/020,079  
; CURRENT FILING DATE: 2001-12-12  
; PRIOR APPLICATION NUMBER: US 60/255,103  
; PRIOR FILING DATE: 2000-12-12  
; PRIOR APPLICATION NUMBER: US 60/289,422  
; PRIOR FILING DATE: 2001-05-08  
; NUMBER OF SEQ ID NOS: 40  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 22  
; LENGTH: 789  
; TYPE: PRT  
; ORGANISM: homo sapiens  
; FEATURE:  
; NAME/KEY: VARIANT  
; LOCATION: (1)...(789)  
; OTHER INFORMATION: Xaa = Any Amino Acid  
US-10-020-079-22

Query Match 56.2%; Score 50; DB 4; Length 789;  
Best Local Similarity 88.9%; Pred. No. 73;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 HQPPQPLPP 10  
| |||||  
Db 610 HLPPQPLPP 618

RESULT 13

US-10-020-079-40

; Sequence 40, Application US/10020079  
; Patent No. 6579710  
; GENERAL INFORMATION:  
; APPLICANT: Turner, C. Alexander Jr.  
; APPLICANT: Mathur, Brian  
; APPLICANT: Friddle, Carl Johan

```
; TITLE OF INVENTION: No. 6579710e1 Human Kinases and Polynucleotides Encoding
the Same
; FILE REFERENCE: LEX-0281-USA
; CURRENT APPLICATION NUMBER: US/10/020,079
; CURRENT FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: US 60/255,103
; PRIOR FILING DATE: 2000-12-12
; PRIOR APPLICATION NUMBER: US 60/289,422
; PRIOR FILING DATE: 2001-05-08
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 40
; LENGTH: 838
; TYPE: PRT
; ORGANISM: homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(838)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-10-020-079-40
```

```
Query Match          56.2%; Score 50; DB 4; Length 838;
Best Local Similarity 88.9%; Pred. No. 77;
Matches      8; Conservative    0; Mismatches    1; Indels      0; Gaps      0;
```

```
Qy          2 HQPPQPLPP 10
              | |||||
Db          684 HLPPQPLPP 692
```

#### RESULT 14

```
US-10-020-079-38
; Sequence 38, Application US/10020079
; Patent No. 6579710
; GENERAL INFORMATION:
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Mathur, Brian
; APPLICANT: Friddle, Carl Johan
; TITLE OF INVENTION: No. 6579710e1 Human Kinases and Polynucleotides Encoding
the Same
; FILE REFERENCE: LEX-0281-USA
; CURRENT APPLICATION NUMBER: US/10/020,079
; CURRENT FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: US 60/255,103
; PRIOR FILING DATE: 2000-12-12
; PRIOR APPLICATION NUMBER: US 60/289,422
; PRIOR FILING DATE: 2001-05-08
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 38
; LENGTH: 851
; TYPE: PRT
; ORGANISM: homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(851)
; OTHER INFORMATION: Xaa = Any Amino Acid
```



US-10-020-079-38

Query Match 56.2%; Score 50; DB 4; Length 851;  
Best Local Similarity 88.9%; Pred. No. 78;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 HQPPQPLPP 10  
| |||||  
Db 697 HLPPQPLPP 705

RESULT 15

US-10-020-079-32

; Sequence 32, Application US/10020079  
; Patent No. 6579710  
; GENERAL INFORMATION:  
; APPLICANT: Turner, C. Alexander Jr.  
; APPLICANT: Mathur, Brian  
; APPLICANT: Friddle, Carl Johan  
; TITLE OF INVENTION: No. 6579710el Human Kinases and Polynucleotides Encoding  
the Same  
; FILE REFERENCE: LEX-0281-USA  
; CURRENT APPLICATION NUMBER: US/10/020,079  
; CURRENT FILING DATE: 2001-12-12  
; PRIOR APPLICATION NUMBER: US 60/255,103  
; PRIOR FILING DATE: 2000-12-12  
; PRIOR APPLICATION NUMBER: US 60/289,422  
; PRIOR FILING DATE: 2001-05-08  
; NUMBER OF SEQ ID NOS: 40  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 32  
; LENGTH: 863  
; TYPE: PRT  
; ORGANISM: homo sapiens  
; FEATURE:  
; NAME/KEY: VARIANT  
; LOCATION: (1)...(863)  
; OTHER INFORMATION: Xaa = Any Amino Acid  
US-10-020-079-32

Query Match 56.2%; Score 50; DB 4; Length 863;  
Best Local Similarity 88.9%; Pred. No. 79;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 HQPPQPLPP 10  
| |||||  
Db 684 HLPPQPLPP 692

Search completed: August 24, 2004, 15:55:22  
Job time : 17.4552 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 24, 2004, 15:26:28 ; Search time 14.5522 Seconds  
(without alignments)  
99.151 Million cell updates/sec

Title: US-09-641-801-34  
Perfect score: 89  
Sequence: 1 MHQPPQPLPPTVMFP 15

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_78:\*  
1: pirl:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	%		DB	ID	Description
		Query	Match Length			
1	89	100.0	222	2	A32979	beta-casein precur
2	81	91.0	222	2	JC1384	beta-casein precur
3	80	89.9	209	2	A59068	beta-casein varian
4	80	89.9	224	1	KBBOA2	beta-casein precur
5	55	61.8	232	2	A48384	beta-casein - pig
6	53	59.6	4957	2	T03455	ALR protein - huma
7	53	59.6	5262	2	T03454	ALR protein - huma
8	50	56.2	64	2	A60305	beta-casein - Arab
9	50	56.2	295	2	T23989	hypothetical prote
10	49	55.1	292	2	T26908	hypothetical prote
11	49	55.1	346	2	T01123	hypothetical prote
12	47	52.8	191	2	E87601	OmpA family protei
13	47	52.8	226	1	KBHU	beta-casein precur

14	47	52.8	270	2	H83619	hypothetical prote
15	47	52.8	429	2	S01919	knirps protein - f
16	47	52.8	479	1	IXBE1F	alpha trans-induci
17	47	52.8	490	1	IXBE17	alpha trans-induci
18	47	52.8	490	1	IXBE33	alpha trans-induci
19	47	52.8	490	2	JS0689	virion protein 16
20	47	52.8	683	2	T34103	hypothetical prote
21	47	52.8	907	2	E96636	hypothetical prote
22	47	52.8	1264	2	A36858	G2R protein - vari
23	47	52.8	1896	2	B72175	D15R protein - var
24	47	52.8	1897	2	T28621	hypothetical prote
25	46	51.7	435	2	T46443	hypothetical prote
26	46	51.7	690	2	S64646	biotin-[acetyl-CoA
27	46	51.7	833	2	T01547	probable phospholi
28	46	51.7	1070	2	S75712	cellulase (EC 3.2.
29	46	51.7	1541	2	T02831	AAA protein L4171.
30	46	51.7	2282	2	T42717	DNA-binding protei
31	45.5	51.1	103	2	JC6167	somatostatin-14 [P
32	45.5	51.1	331	2	T32662	hypothetical prote
33	45	50.6	129	2	G22845	hypothetical prote
34	45	50.6	181	2	G96715	transcription fact
35	45	50.6	283	2	T49546	hypothetical prote
36	45	50.6	315	2	T49982	hypothetical prote
37	45	50.6	330	2	T05717	probable extensin
38	45	50.6	416	2	S27198	homeotic protein H
39	45	50.6	427	2	T48159	hypothetical prote
40	45	50.6	503	2	T05347	hypothetical prote
41	45	50.6	998	2	T30930	hypothetical prote
42	45	50.6	1274	2	T37193	enamelin matrix pr
43	45	50.6	1316	2	T00381	KIAA0633 protein -
44	45	50.6	2414	2	A54277	transcription adap
45	44.5	50.0	505	2	A53152	annexin XI - human

## ALIGNMENTS

### RESULT 1

A32979

beta-casein precursor - sheep

C;Species: *Ovis orientalis aries*, *Ovis ammon aries* (domestic sheep)

C;Date: 12-Oct-1989 #sequence\_revision 31-Dec-1993 #text\_change 13-Aug-1999

C;Accession: A32979; A29173

R;Provot, C.; Persuy, M.A.; Mercier, J.C.

Biochimie 71, 827-832, 1989

A;Title: Complete nucleotide sequence of ovine beta-casein cDNA: inter-species comparison.

A;Reference number: A32979; MUID:89375530; PMID:2505862

A;Accession: A32979

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-222 <PRO>

A;Cross-references: GB:X16482; NID:g1210; PIDN:CAA34502.1; PID:g1211

A;Note: the authors translated the codon CAC for residue 160 as Lys

R;Richardson, B.C.; Mercier, J.C.

Eur. J. Biochem. 99, 285-297, 1979

A;Title: The primary structure of the ovine beta-caseins.

A;Reference number: A29173; MUID:80046695; PMID:499202  
A;Accession: A29173  
A;Molecule type: protein  
A;Residues: 16-69,'T',71-77,'P',79-81,'A',83-222 <RIC>  
C;Superfamily: beta-casein  
C;Keywords: milk; phosphoprotein  
F;1-15/Domain: signal sequence #status predicted <SIG>  
F;16-222/Product: beta-casein #status experimental <MAT>  
F;30,32,33,34,50/Binding site: phosphate (Ser) (covalent) (by casein kinase II)  
#status predicted

Query Match 100.0%; Score 89; DB 2; Length 222;  
Best Local Similarity 100.0%; Pred. No. 7e-05;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MHQPPQQLPPTVMFP 15  
|||||||  
Db 159 MHQPPQQLPPTVMFP 173

#### RESULT 2

JC1384

beta-casein precursor - goat

C;Species: Capra aegagrus hircus (domestic goat)

C;Date: 10-Jun-1993 #sequence\_revision 10-Jun-1993 #text\_change 23-Feb-1997

C;Accession: JC1384

R;Roberts, B.; DiTullio, P.; Vitale, J.; Hehir, K.; Gordon, K.

Gene 121, 255-262, 1992

A;Title: Cloning of the goat beta-casein-encoding gene and expression in transgenic mice.

A;Reference number: JC1384; MUID:93077039; PMID:1446822

A;Accession: JC1384

A;Molecule type: DNA

A;Residues: 1-222 <ROB>

A;Cross-references: GB:M90556

C;Genetics:

A;Gene: CSN2

A;Introns: 17/3; 26/3; 35/3; 45/3; 57/3; 221/3

C;Superfamily: beta-casein

C;Keywords: milk; phosphoprotein

Query Match 91.0%; Score 81; DB 2; Length 222;  
Best Local Similarity 93.3%; Pred. No. 0.00079;  
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MHQPPQQLPPTVMFP 15  
||||||| |||||  
Db 159 MHQPPQQLSPTVMFP 173

#### RESULT 3

A59068

beta-casein variant CnH - bovine

C;Species: Bos primigenius taurus (cattle)

C;Date: 20-Sep-1999 #sequence\_revision 24-Sep-1999 #text\_change 24-Sep-1999

C;Accession: A59068; B59068

R;Han, S.K.; Shin, Y.C.

Anim. Genet. 27(Suppl.2), 91b, 1996  
 A;Title: Biochemical characterization of the new beta-casein variant in Korean cattle.  
 A;Reference number: A59068  
 A;Accession: A59068  
 A;Status: protein sequence not shown  
 A;Molecule type: protein  
 A;Residues: 1-209 <HAN1>  
 A;Experimental source: strain Korean cattle  
 A;Note: submitted to the Protein Sequence Database, September 1999  
 A;Note: includes casein phosphopeptide H  
 A;Accession: B59068  
 A;Status: protein sequence not shown  
 A;Molecule type: protein  
 A;Residues: 1-28 <HAN2>  
 A;Experimental source: strain Korean cattle  
 C;Superfamily: beta-casein  
 C;Keywords: milk; phosphoprotein  
 F;15,17,18,19/Binding site: phosphate (Ser) (covalent) #status predicted

Query Match 89.9%; Score 80; DB 2; Length 209;  
 Best Local Similarity 93.3%; Pred. No. 0.001;  
 Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MHQPPQPLPPTVMFP 15  
 |||| |||||  
 Db 144 MHQPHQPLPPTVMFP 158

#### RESULT 4

KBBOA2

beta-casein precursor - bovine

C;Species: Bos primigenius taurus (cattle)

C;Date: 24-Apr-1984 #sequence\_revision 12-May-1995 #text\_change 11-May-2000

C;Accession: I45873; B29087; S01860; A25846; S02429; A90489; A91191; B91192; C91192; D91192; A90739; I46963; A91413; A03110

R;Bonsing, J.; Ring, J.M.; Stewart, A.F.; Mackinlay, A.G.

Aust. J. Biol. Sci. 41, 527-537, 1988

A;Title: Complete nucleotide sequence of the bovine beta-casein gene.

A;Reference number: I45873; MUID:90147279; PMID:3271384

A;Accession: I45873

A;Status: translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-81, 'H', 83-224 <BON>

A;Cross-references: GB:M55158; NID:g162804; PIDN:AAA30431.1; PID:g162805

R;Stewart, A.F.; Bonsing, J.; Beattie, C.W.; Shah, F.; Willis, I.M.; Mackinlay, A.G.

Mol. Biol. Evol. 4, 231-241, 1987

A;Title: Complete nucleotide sequences of bovine alpha-s2- and beta-casein

cDNAs: comparisons with related sequences in other species.

A;Reference number: A93062; MUID:88188989; PMID:2833669

A;Accession: B29087

A;Status: translation not shown

A;Molecule type: mRNA

A;Residues: 1-224 <STE>

A;Cross-references: GB:M16645; NID:g162930; PIDN:AAA30480.1; PID:g162931

A;Experimental source: A2 variant

R;Baev, A.A.; Smirnov, I.K.; Gorodetskii, S.I.  
 Mol. Biol. 21, 214-222, 1987  
 A;Title: Primary structure of bovine beta-casein cDNA.  
 A;Reference number: S01860  
 A;Accession: S01860  
 A;Molecule type: mRNA  
 A;Residues: 1-81,'H',83-224 <BAE>  
 A;Cross-references: EMBL:X06359; NID:g171; PIDN:CAA29658.1; PID:g757752  
 A;Experimental source: A1 variant  
 A;Note: this paper is a translation of the Russian paper published in Mol. Biol. Moscow (1987) 21: 255-265  
 R;Jimenez-Flores, R.; Kang, Y.C.; Richardson, T.  
 Biochem. Biophys. Res. Commun. 142, 617-621, 1987  
 A;Title: Cloning and sequence analysis of bovine beta-casein cDNA.  
 A;Reference number: A25846; MUID:87128158; PMID:3814153  
 A;Accession: A25846  
 A;Molecule type: mRNA  
 A;Residues: 1-107,'L',109-151,'PL',154-209,'Q',211-224 <JIM>  
 A;Cross-references: GB:M15132; NID:g162796; PIDN:AAA30430.1; PID:g162797  
 R;Carles, C.; Huet, J.C.; Ribadeau-Dumas, B.  
 FEBS Lett. 229, 265-272, 1988  
 A;Title: A new strategy for primary structure determination of proteins: application to bovine beta-casein.  
 A;Reference number: S02429; MUID:88152252; PMID:3278933  
 A;Accession: S02429  
 A;Molecule type: protein  
 A;Residues: 16-81,'H',83-224 <CAR>  
 A;Experimental source: A1 variant  
 R;Yan, S.B.; Wold, F.  
 Biochemistry 23, 3759-3765, 1984  
 A;Title: Neoglycoproteins: in vitro introduction of glycosyl units at glutaminesin beta-casein using transglutaminase.  
 A;Reference number: A90489; MUID:85000478; PMID:6148101  
 A;Accession: A90489  
 A;Molecule type: protein  
 A;Residues: 16-224 <YAN>  
 R;Ribadeau-Dumas, B.; Brignon, G.; Grosclaude, F.; Mercier, J.C.  
 Eur. J. Biochem. 25, 505-514, 1972  
 A;Title: Structure primaire de la caseine beta bovine.  
 A;Reference number: A91191; MUID:72233212; PMID:4557764  
 A;Accession: A91191  
 A;Molecule type: protein  
 A;Residues: 16-131,'Q',133-151,'PL',154-189,'E',191-209,'Q',211-224 <RIB>  
 A;Experimental source: A2 variant  
 A;Note: article in French with an English abstract  
 R;Grosclaude, F.; Mahe, M.F.; Mercier, J.C.; Ribadeau-Dumas, B.  
 Eur. J. Biochem. 26, 328-337, 1972  
 A;Title: Caracterisation des variants genetiques des caseines alpha-S1 et beta bovines.  
 A;Reference number: A91192; MUID:72214259; PMID:5064450  
 A;Note: article in French with an English abstract  
 A;Accession: B91192  
 A;Molecule type: protein  
 A;Residues: 16-81,'H',83-131,'Q',133-151,'PL',154-189,'E',191-209,'Q',211-224 <VA1>  
 A;Experimental source: A1 variant  
 A;Accession: C91192

A;Molecule type: protein  
 A;Residues: 16-81,'H',83-131,'Q',133-136,'R',138-151,'PL',154-189,'E',191-209,'Q',211-224 <VAB>  
 A;Experimental source: B variant  
 A;Accession: D91192  
 A;Molecule type: protein  
 A;Residues: 16-51,'K',53-81,'H',83-131,'Q',133-151,'PL',154-189,'E',191-209,'Q',211-224 <VAC>  
 A;Experimental source: C variant  
 A;Note: this variant lacks a phosphate group on 50-Ser  
 R;Ribadeau-Dumas, B.; Grosclaude, F.; Mercier, J.C.  
 C. R. Acad. Sci. Hebd. Seances Acad. Sci. D 270, 2369-2372, 1970  
 A;Title: Localisation dans la chaine peptidique de la caseine beta bovine de la substitution His/Gln differenciant les variants genetiques A2 et A3.  
 A;Reference number: A90739; MUID:71252171; PMID:4997616  
 A;Note: article in French with an English abstract  
 A;Accession: A90739  
 A;Molecule type: protein  
 A;Residues: 118-120,'Q',122-124 <VA3>  
 A;Experimental source: A3 variant  
 R;Simons, G.; van den Heuvel, W.; Reynen, T.; Frijters, A.; Rutten, G.; Slangen, C.J.; Groenen, M.; de Vos, W.M.; Siezen, R.J.  
 Protein Eng. 6, 763-770, 1993  
 A;Title: Overproduction of bovine beta-casein in Escherichia coli and engineering of its main chymosin cleavage site.  
 A;Reference number: I46963; MUID:94068382; PMID:8248100  
 A;Accession: I46963  
 A;Status: translated from GB/EMBL/DDBJ  
 A;Molecule type: mRNA  
 A;Residues: 1-120,'Q',122-224 <SIM>  
 A;Cross-references: GB:S67277; NID:g459291; PIDN:AAB29137.1; PID:g459292  
 A;Experimental source: A3 variant  
 R;Grosclaude, F.; Mahe, M.F.; Voglino, G.F.  
 FEBS Lett. 45, 3-5, 1974  
 A;Title: Le variant beta-E et le code de phosphorylation des caseines bovines.  
 A;Reference number: A91413; MUID:75005247; PMID:4411121  
 A;Note: article in French with an English abstract  
 A;Accession: A91413  
 A;Molecule type: protein  
 A;Residues: 48-50,'K',52-63 <VAE>  
 A;Experimental source: E variant  
 A;Note: 50-Ser is phosphorylated  
 C;Comment: The sequence shown is the A2 variant.  
 C;Genetics:  
 A;Introns: 17/3; 26/3; 35/3; 43/3; 57/3; 223/3  
 C;Superfamily: beta-casein  
 C;Keywords: milk; phosphoprotein  
 F;1-15/Domain: signal sequence #status predicted <SIG>  
 F;16-224/Product: beta-casein #status experimental <MAT>  
 F;30,32,33,34/Binding site: phosphate (Ser) (covalent) (by casein kinase II) #status experimental  
 F;50/Binding site: phosphate (Ser) (covalent) (by casein kinase II) (partial) #status experimental

Query Match 89.9%; Score 80; DB 1; Length 224;  
 Best Local Similarity 93.3%; Pred. No. 0.0011;  
 Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MHQPPQPLPPTVMFP 15  
||||| |||||  
Db 159 MHQPHQPLPPTVMFP 173

RESULT 5

A48384

beta-casein - pig

C;Species: *Sus scrofa domestica* (domestic pig)

C;Date: 19-Nov-1993 #sequence\_revision 18-Nov-1994 #text\_change 03-May-1996

C;Accession: A48384

R;Alexander, L.J.; Beattie, C.W.

Anim. Genet. 23, 369-371, 1992

A;Title: The sequence of porcine beta-casein cDNA.

A;Reference number: A48384; MUID:92367961; PMID:1503277

A;Accession: A48384

A;Status: preliminary

A;Molecule type: nucleic acid

A;Residues: 1-232 <ALE>

A;Experimental source: mammary gland

A;Note: sequence inconsistent with the nucleotide translation

A;Note: sequence extracted from NCBI backbone (NCBIN:110891, NCBIP:110895)

C;Superfamily: beta-casein

Query Match 61.8%; Score 55; DB 2; Length 232;  
Best Local Similarity 71.4%; Pred. No. 2.2;  
Matches 10; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 MHQPPQPLPPTVMF 14  
||| |||:| | ||  
Db 158 MHQIPQVPQTPMF 171

RESULT 6

T03455

ALR protein - human

C;Species: *Homo sapiens* (man)

C;Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 27-Oct-2003

C;Accession: T03455

R;Prasad, R.; Zhadanov, A.B.; Sedkov, Y.; Bullrich, F.; Druck, T.; Rallapalli, R.; Yano, T.; Alder, H.; Croce, C.M.; Huebner, K.; Mazo, A.; Canaani, E.  
Oncogene 15, 549-560, 1997

A;Title: Structure and expression pattern of human ALR, a novel gene with strong homology to ALL-1 involved in acute leukemia, and to *Drosophila trithorax*.

A;Reference number: Z14954; MUID:97388474; PMID:9247308

A;Accession: T03455

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: mRNA

A;Residues: 1-4957 <PRA>

A;Cross-references: EMBL:AF010404; NID:g2358286; PIDN:AAC51735.1; PID:g2358287

C;Genetics:

A;Gene: ALR

A;Map position: 12

C;Superfamily: acute lymphoblastic leukemia protein, ALR type

C;Keywords: alternative splicing



Query Match 59.6%; Score 53; DB 2; Length 4957;  
Best Local Similarity 57.1%; Pred. No. 86;  
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 MHQPPQPLPPTVMF 14  
:|:|:| | | | |  
Db 1886 LHKPPRPQPPEVAF 1899

RESULT 7

T03454

ALR protein - human

C;Species: Homo sapiens (man)

C;Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 27-Oct-2003

C;Accession: T03454

R;Prasad, R.; Zhadanov, A.B.; Sedkov, Y.; Bullrich, F.; Druck, T.; Rallapalli, R.; Yano, T.; Alder, H.; Croce, C.M.; Huebner, K.; Mazo, A.; Canaani, E.  
Oncogene 15, 549-560, 1997

A;Title: Structure and expression pattern of human ALR, a novel gene with strong homology to ALL-1 involved in acute leukemia, and to Drosophila trithorax.

A;Reference number: Z14954; MUID:97388474; PMID:9247308

A;Accession: T03454

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: mRNA

A;Residues: 1-5262 <PRA>

A;Cross-references: EMBL:AF010403; NID:g2358284; PIDN:AAC51734.1; PID:g2358285

C;Genetics:

A;Gene: ALR

A;Map position: 12

C;Superfamily: acute lymphoblastic leukemia protein, ALR type

C;Keywords: alternative splicing

Query Match 59.6%; Score 53; DB 2; Length 5262;  
Best Local Similarity 57.1%; Pred. No. 91;  
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 MHQPPQPLPPTVMF 14  
:|:|:| | | | |  
Db 2191 LHKPPRPQPPEVAF 2204

RESULT 8

A60305

beta-casein - Arabian camel (fragment)

C;Species: Camelus dromedarius (Arabian camel)

C;Date: 10-Nov-1992 #sequence\_revision 10-Nov-1992 #text\_change 30-Sep-1993

C;Accession: A60305

R;Beg, O.U.; von Bahr-Lindstroem, H.; Zaidi, Z.H.; Joernvall, H.  
Regul. Pept. 15, 55-62, 1986

A;Title: Characterization of a camel milk protein rich in proline identifies a new beta-casein fragment.

A;Reference number: A60305; MUID:87017861; PMID:3763959

A;Accession: A60305

A;Molecule type: protein

A;Residues: 1-64 <BEG>

A;Note: this fragment appears to form by a non-tryptic cleavage of beta-casein in camel milk

C;Superfamily: beta-casein

Query Match 56.2%; Score 50; DB 2; Length 64;  
Best Local Similarity 60.0%; Pred. No. 2.8;  
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 1 MHQPPQPLPPTVMFP 15  
|:| |||:| | | |  
Db 1 MYQIPQVPQTPMIP 15

RESULT 9

T23989

hypothetical protein R07A4.3 - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 18-Feb-2000

C;Accession: T23989

R;Cottage, A.

submitted to the EMBL Data Library, November 1995

A;Reference number: Z19827

A;Accession: T23989

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-295 <WIL>

A;Cross-references: EMBL:Z67756; PIDN:CAA91763.1; GSPDB:GN00028; CESP:R07A4.3

A;Experimental source: clone R07A4

C;Genetics:

A;Gene: CESP:R07A4.3

A;Map position: X

A;Introns: 29/3; 53/1; 142/1; 168/1; 229/1; 253/1

Query Match 56.2%; Score 50; DB 2; Length 295;  
Best Local Similarity 75.0%; Pred. No. 13;  
Matches 9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 4 PPQPLPPTVMFP 15  
||||| |||: |  
Db 102 PPQPLKPTVIRP 113

RESULT 10

T26908

hypothetical protein Y45F10A.1 - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 04-Mar-2000

C;Accession: T26908

R;McMurray, A.

submitted to the EMBL Data Library, January 1998

A;Reference number: Z20285

A;Accession: T26908

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-292 <WIL>

A;Cross-references: EMBL:AL021488; PIDN:CAA16365.1; GSPDB:GN00022;

CESP:Y45F10A.1

A;Experimental source: clone Y45F10A

C;Genetics:

A;Gene: CESP:Y45F10A.1

A;Map position: 4

A;Introns: 228/2; 261/3

C;Superfamily: Caenorhabditis elegans hypothetical protein Y45F10A.1

Query Match 55.1%; Score 49; DB 2; Length 292;  
Best Local Similarity 60.0%; Pred. No. 17;  
Matches 9; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 MHQPPQPLPPTVMFP 15

|| | |||| |

Db 1 MHSPNHPLPPTSNSP 15

# RESULT 11

T01123

hypothetical protein At2g32840 [imported] - Arabidopsis thaliana

N;Alternate names: hypothetical protein F24L7.2; T21L14.22

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 16-Feb-2001

C;Accession: T01123; T00784; B84738

R;Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul, S.; Mason, T.M.; Kerlavage, A.R.; Adams, M.D.; Somerville, C.R.; Venter, J.C.

submitted to the EMBL Data Library, December 1997

A;Description: Arabidopsis thaliana chromosome II BAC T21L14 genomic sequence.

A;Reference number: Z14209

A;Accession: T01123

A;Status: translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-346 <ROU>

A;Cross-references: EMBL:AC003033; NID:g2702261; PID:g2702278

A;Experimental source: cultivar Columbia

R;Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul, S.; Mason, T.M.; Kerlavage, A.R.; Adams, M.D.; Somerville, C.R.; Venter, J.C.

submitted to the EMBL Data Library, February 1998

A;Description: Arabidopsis thaliana chromosome II BAC F24L7 genomic sequence.

A;Reference number: Z14204

A;Accession: T00784

A;Status: translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-346 <ROW>

A;Cross-references: EMBL:AC003974; NID:g2914688; PID:g2914690

A;Experimental source: cultivar Columbia

R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; Mason, T.M.; Bowman, C.L.; Barnstead, M.E.; Feldblyum, T.V.; Buell, C.R.; Ketchum, K.A.; Lee, J.J.; Ronning, C.M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.J.; Gill, J.E.; Adams, M.D.; Carrera, A.J.; Creasy, T.H.; Goodman, H.M.; Somerville, C.R.; Copenhaver, G.P.; Preuss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.C.

Nature 402, 761-768, 1999

A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A;Reference number: A84420; MUID:20083487; PMID:10617197

A;Accession: B84738

A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-346 <STO>  
A;Cross-references: GB:AE002093; NID:g2702278; PIDN:AAB91981.1; GSPDB:GN00139  
C;Genetics:  
A;Gene: At2g32840; T21L14.22; F24L7.2  
A;Map position: 2  
A;Introns: 185/3; 196/3; 214/2; 227/3; 253/3; 289/3; 316/2

Query Match 55.1%; Score 49; DB 2; Length 346;  
Best Local Similarity 50.0%; Pred. No. 20;  
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Qy 2 HQPPQPLPPTVMFP 15  
| | | | | : : : |  
Db 87 HQPPHPDPSSLIYP 100

RESULT 12

E87601

OmpA family protein [imported] - *Caulobacter crescentus*

C;Species: *Caulobacter crescentus*

C;Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 20-Apr-2001

C;Accession: E87601

R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.;  
Heidelberg, J.F.; Alley, M.; Ohta, N.; Maddock, J.R.; Potocka, I.; Nelson, W.C.;  
Newton, A.; Stephens, C.; Phadke, N.D.; Ely, B.; Laub, M.T.; DeBoy, R.T.;  
Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolonay, J.F.; Smit, J.;  
Craven, M.; Khouri, H.; Shetty, J.; Berry, K.; Utterback, T.; Tran, K.; Wolf,  
A.; Vamathevan, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.;  
Venter, J.C.; Fraser, C.M.

Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A;Title: Complete Genome Sequence of *Caulobacter crescentus*.

A;Reference number: A87249; MUID:21173698; PMID:11259647

A;Accession: E87601

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-191 <STO>

A;Cross-references: GB:AE005673; NID:g13424457; PIDN:AAK24809.1; GSPDB:GN00148

C;Genetics:

A;Gene: CC2845

Query Match 52.8%; Score 47; DB 2; Length 191;  
Best Local Similarity 66.7%; Pred. No. 21;  
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 4 PPQPLPPTVMFP 15  
| | | | | : |  
Db 60 PPQPLPPAPLPP 71

RESULT 13

KBHU

beta-casein precursor [validated] - human

C;Species: *Homo sapiens* (man)

C;Date: 30-Jun-1988 #sequence\_revision 15-Aug-1997 #text\_change 08-Dec-2000

C;Accession: I53730; S08040; S04049; S11072; A27219; A30773

R;Hansson, L.; Edlund, A.; Johansson, T.; Hernell, O.; Stromqvist, M.; Lindquist, S.; Lonnerdal, B.; Bergstrom, S.  
Gene 139, 193-199, 1994  
A;Title: Structure of the human beta-casein encoding gene.  
A;Reference number: I53730; MUID:94156198; PMID:8112603  
A;Accession: I53730  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 1-226 <RES>  
A;Cross-references: GB:L10615; NID:g2695660; PIDN:AAC82978.1; PID:g2695661  
R;Menon, R.S.  
submitted to the EMBL Data Library, October 1989  
A;Reference number: S08040  
A;Accession: S08040  
A;Molecule type: mRNA  
A;Residues: 1-132,'V',134-139,'Q',141-226 <MEN>  
A;Cross-references: EMBL:X17070  
R;Menon, R.S.; Ham, R.G.  
Nucleic Acids Res. 17, 2869, 1989  
A;Title: Human beta-casein: partial cDNA sequence and apparent polymorphism.  
A;Reference number: S04049; MUID:89240053; PMID:2717418  
A;Accession: S04049  
A;Molecule type: mRNA  
A;Residues: 161-226 <MEN2>  
A;Cross-references: EMBL:X13766; NID:g29673; PIDN:CAA32017.1; PID:g29674  
R;Loennerdal, B.; Bergstroem, S.; Andersson, Y.; Hjalmarsson, K.; Sundqvist, A.K.; Hernell, O.  
FEBS Lett. 269, 153-156, 1990  
A;Title: Cloning and sequencing of a cDNA encoding human milk beta-casein.  
A;Reference number: S11072; MUID:90353560; PMID:2387396  
A;Accession: S11072  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-33,35-226 <LOE>  
A;Cross-references: GB:X55739; NID:g288097; PIDN:CAA39270.1; PID:g288098  
R;Greenberg, R.; Groves, M.L.; Dower, H.J.  
J. Biol. Chem. 259, 5132-5138, 1984  
A;Title: Human beta-casein: amino acid sequence and identification of phosphorylation sites.  
A;Reference number: A27219; MUID:84185624; PMID:6715339  
A;Accession: A27219  
A;Molecule type: protein  
A;Residues: 16-29,'P',31-47,'T',49,'Q',51-119,'Q',121-148,'S',150-172,'E',174-181,'E',183,'L',185-187,'V',189-206,'P',208-213,'PE',216,'STTZABH',223-226 <GRE>  
C;Genetics:  
A;Gene: GDB:CSN2; CASB  
A;Cross-references: GDB:125234; OMIM:115460  
A;Map position: 4q13-4q21  
A;Introns: 17/3; 26/3; 33/3; 48/3; 225/3  
C;Superfamily: beta-casein  
C;Keywords: calcium; milk; phosphoprotein  
F;1-15/Domain: signal sequence #status predicted <SIG>  
F;16-226/Product: beta-casein #status experimental <MAT>  
F;18/Binding site: phosphate (Thr) (covalent) #status experimental  
F;21,23,24,25/Binding site: phosphate (Ser) (covalent) #status experimental

Query Match

52.8%; Score 47; DB 1; Length 226;

Best Local Similarity 53.3%; Pred. No. 25;  
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 1 MHQPPQPLPPTVMFP 15  
| | |||:| |: |  
Db 150 MQQVPQPIPQTLALP 164

RESULT 14

H83619

hypothetical protein PA0197 [imported] - *Pseudomonas aeruginosa* (strain PA01)

C;Species: *Pseudomonas aeruginosa*

C;Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000

C;Accession: H83619

R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Brinkman, F.S.L.; Hufnagle, W.O.; Kowalik, D.J.; Lagrou, M.; Garber, R.L.; Goltry, L.; Tolentino, E.; Westbrook-Wadman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim, R.M.; Smith, K.A.; Spencer, D.H.; Wong, G.K.S.; Wu, Z.; Paulsen, I.T.; Reizer, J.; Saier, M.H.; Hancock, R.E.W.; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A;Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen.

A;Reference number: A82950; MUID:20437337; PMID:10984043

A;Accession: H83619

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-270 <STO>

A;Cross-references: GB:AE004458; GB:AE004091; NID:g9946031; PIDN:AAG03586.1; GSPDB:GN00131; PASP:PA0197

A;Experimental source: strain PA01

C;Genetics:

A;Gene: PA0197

C;Superfamily: tonB protein

Query Match 52.8%; Score 47; DB 2; Length 270;  
Best Local Similarity 66.7%; Pred. No. 29;  
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 4 PPQPLPPTVMFP 15  
||:|||| | |  
Db 103 PPEPLPPVVEEP 114

RESULT 15

S01919

knirps protein - fruit fly (*Drosophila melanogaster*)

C;Species: *Drosophila melanogaster*

C;Date: 31-Dec-1990 #sequence\_revision 31-Dec-1990 #text\_change 24-Sep-1998

C;Accession: S01919; S02057

R;Nauber, U.; Pankratz, M.J.; Kienlin, A.; Seifert, E.; Klemm, U.; Jaeckle, H.  
Nature 336, 489-492, 1988

A;Title: Abdominal segmentation of the *Drosophila* embryo requires a hormone receptor-like protein encoded by the gap gene knirps.

A;Reference number: S01919; MUID:89057148; PMID:2904128

A;Accession: S01919

A;Molecule type: DNA

A;Residues: 1-429 <NAU1>  
A;Cross-references: EMBL:X13331  
R;Nauber, U.  
submitted to the EMBL Data Library, October 1988  
A;Reference number: S02057  
A;Accession: S02057  
A;Molecule type: DNA  
A;Residues: 1-106,'L',108-429 <NAU2>  
A;Cross-references: EMBL:X13331; NID:g8153; PID:g8154  
C;Genetics:  
A;Gene: knirps  
A;Cross-references: FlyBase:FBgn0001320  
A;Introns: 26/3  
C;Keywords: DNA binding; nucleus; transcription regulation; zinc finger

Query Match 52.8%; Score 47; DB 2; Length 429;  
Best Local Similarity 57.1%; Pred. No. 46;  
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 2 HQPPQPLPPTVMFP 15  
|| | ||| ::||  
Db 183 HQSPFQLPPHLLFP 196

Search completed: August 24, 2004, 15:53:03  
Job time : 16.5522 secs

OM protein - protein search, using sw model

Run on: August 24, 2004, 15:51:19 ; Search time 54.291 Seconds  
 (without alignments)  
 86.825 Million cell updates/sec

Title: US-09-641-801-34  
 Perfect score: 89  
 Sequence: 1 MHQPPQPLPPTVMFP 15

Scoring table: BLOSUM62  
 Gapop 10.0 , Gapext 0.5

Searched: 1295152 seqs, 314255058 residues

Total number of hits satisfying chosen parameters: 1295152

Minimum DB seq length: 0  
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
 Maximum Match 100%  
 Listing first 45 summaries

Database : Published Applications\_AA:\*  
 1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*  
 2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep:\*  
 3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep:\*  
 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep:\*  
 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep:\*  
 6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep:\*  
 7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep:\*  
 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep:\*  
 9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep:\*  
 10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep:\*  
 11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep:\*  
 12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep:\*  
 13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep:\*  
 14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep:\*  
 15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep:\*  
 16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep:\*  
 17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep:\*  
 18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
 score greater than or equal to the score of the result being printed,  
 and is derived by analysis of the total score distribution.

# SUMMARIES

Result	Score	Match	Length	DB	ID	Description
No.						



1	89	100.0	15	14	US-10-281-652-34	Sequence 34, Appl
2	57	64.0	10	14	US-10-281-652-25	Sequence 25, Appl
3	56	62.9	412	16	US-10-408-765A-1080	Sequence 1080, Ap
4	55	61.8	248	16	US-10-437-963-152994	Sequence 152994,
5	53	59.6	10	9	US-09-794-346-1	Sequence 1, Appli
6	53	59.6	10	16	US-10-380-371-1	Sequence 1, Appli
7	53	59.6	60	9	US-09-867-550-978	Sequence 978, App
8	53	59.6	2683	16	US-10-437-963-185370	Sequence 185370,
9	53	59.6	4952	15	US-10-051-874-56	Sequence 56, Appl
10	53	59.6	5008	15	US-10-051-874-166	Sequence 166, App
11	53	59.6	5159	15	US-10-085-198-112	Sequence 112, App
12	53	59.6	5262	15	US-10-051-874-165	Sequence 165, App
13	53	59.6	5262	15	US-10-051-874-167	Sequence 167, App
14	52	58.4	643	16	US-10-437-963-162903	Sequence 162903,
15	51	57.3	95	16	US-10-437-963-161687	Sequence 161687,
16	51	57.3	1303	16	US-10-437-963-195407	Sequence 195407,
17	51	57.3	2771	16	US-10-437-963-195406	Sequence 195406,
18	50	56.2	55	12	US-10-424-599-266218	Sequence 266218,
19	50	56.2	327	14	US-10-017-161-2258	Sequence 2258, Ap
20	50	56.2	327	15	US-10-292-798-1904	Sequence 1904, Ap
21	50	56.2	751	12	US-10-413-437-8	Sequence 8, Appli
22	50	56.2	751	13	US-10-020-079-8	Sequence 8, Appli
23	50	56.2	764	12	US-10-413-437-6	Sequence 6, Appli
24	50	56.2	764	13	US-10-020-079-6	Sequence 6, Appli
25	50	56.2	776	12	US-10-413-437-24	Sequence 24, Appl
26	50	56.2	776	13	US-10-020-079-24	Sequence 24, Appl
27	50	56.2	789	12	US-10-413-437-22	Sequence 22, Appl
28	50	56.2	789	13	US-10-020-079-22	Sequence 22, Appl
29	50	56.2	838	12	US-10-413-437-40	Sequence 40, Appl
30	50	56.2	838	13	US-10-020-079-40	Sequence 40, Appl
31	50	56.2	851	12	US-10-413-437-38	Sequence 38, Appl
32	50	56.2	851	13	US-10-020-079-38	Sequence 38, Appl
33	50	56.2	863	12	US-10-413-437-32	Sequence 32, Appl
34	50	56.2	863	13	US-10-020-079-32	Sequence 32, Appl
35	50	56.2	864	12	US-10-413-437-4	Sequence 4, Appli
36	50	56.2	864	13	US-10-020-079-4	Sequence 4, Appli
37	50	56.2	870	12	US-10-413-437-2	Sequence 2, Appli
38	50	56.2	870	13	US-10-020-079-2	Sequence 2, Appli
39	50	56.2	876	12	US-10-413-437-30	Sequence 30, Appl
40	50	56.2	876	13	US-10-020-079-30	Sequence 30, Appl
41	50	56.2	889	12	US-10-413-437-20	Sequence 20, Appl
42	50	56.2	889	13	US-10-020-079-20	Sequence 20, Appl
43	50	56.2	895	12	US-10-413-437-18	Sequence 18, Appl
44	50	56.2	895	13	US-10-020-079-18	Sequence 18, Appl
45	50	56.2	951	12	US-10-413-437-36	Sequence 36, Appl

#### ALIGNMENTS

RESULT 1

US-10-281-652-34

; Sequence 34, Application US/10281652

; Publication No. US20030091606A1

; GENERAL INFORMATION:

; APPLICANT: STANTON, G. John

```
; APPLICANT: HUGHES, Thomas K.
; APPLICANT: BOLDOGH, Istvan
; TITLE OF INVENTION: USE OF COLOSTRININ, CONSTITUENT PEPTIDES THEREOF, AND
; TITLE OF INVENTION: ANALOGS THEREOF AS OXIDATIVE STRESS REGULATORS
; FILE REFERENCE: 265.00220101
; CURRENT APPLICATION NUMBER: US/10/281,652
; CURRENT FILING DATE: 2002-10-28
; PRIOR APPLICATION NUMBER: US/09/641,803
; PRIOR FILING DATE: 2000-08-17
; PRIOR APPLICATION NUMBER: 60/149,310
; PRIOR FILING DATE: 1999-08-17
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 34
; LENGTH: 15
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: peptide
US-10-281-652-34
```

```
Query Match          100.0%; Score 89; DB 14; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.0004;
Matches    15; Conservative    0; Mismatches    0; Indels    0; Gaps    0;
```

```
Qy      1 MHQPPQPLPPTVMFP 15
          |||||
Db      1 MHQPPQPLPPTVMFP 15
```

# RESULT 2

US-10-281-652-25

```
; Sequence 25, Application US/10281652
; Publication No. US20030091606A1
; GENERAL INFORMATION:
; APPLICANT: STANTON, G. John
; APPLICANT: HUGHES, Thomas K.
; APPLICANT: BOLDOGH, Istvan
; TITLE OF INVENTION: USE OF COLOSTRININ, CONSTITUENT PEPTIDES THEREOF, AND
; TITLE OF INVENTION: ANALOGS THEREOF AS OXIDATIVE STRESS REGULATORS
; FILE REFERENCE: 265.00220101
; CURRENT APPLICATION NUMBER: US/10/281,652
; CURRENT FILING DATE: 2002-10-28
; PRIOR APPLICATION NUMBER: US/09/641,803
; PRIOR FILING DATE: 2000-08-17
; PRIOR APPLICATION NUMBER: 60/149,310
; PRIOR FILING DATE: 1999-08-17
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 25
; LENGTH: 10
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: peptide
```

US-10-281-652-25

Query Match 64.0%; Score 57; DB 14; Length 10;  
Best Local Similarity 100.0%; Pred. No. 1.5;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 QPLPPTVMFP 15  
| | | | | | | |  
Db 1 QPLPPTVMFP 10

RESULT 3

US-10-408-765A-1080

; Sequence 1080, Application US/10408765A  
; Publication No. US20040101874A1  
; GENERAL INFORMATION:  
; APPLICANT: Ghosh, Soumitra S.  
; APPLICANT: Fahy, Eoin D.  
; APPLICANT: Zhang, Bing  
; APPLICANT: Gibson, Bradford W.  
; APPLICANT: Taylor, Steven W.  
; APPLICANT: Glenn, Gary M.  
; APPLICANT: Warnock, Dale E.  
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION  
; TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME  
; FILE REFERENCE: 660088.465  
; CURRENT APPLICATION NUMBER: US/10/408,765A  
; CURRENT FILING DATE: 2003-04-04  
; NUMBER OF SEQ ID NOS: 3077  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 1080  
; LENGTH: 412  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-408-765A-1080

Query Match 62.9%; Score 56; DB 16; Length 412;  
Best Local Similarity 69.2%; Pred. No. 57;  
Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 3 QPPQPLPPTVMFP 15  
| | | | | | | : | |  
Db 15 QPPQPAPPPPLFP 27

RESULT 4

US-10-437-963-152994

; Sequence 152994, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Barbazuk, Brad

```
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules
Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 152994
; LENGTH: 248
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(248)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_52991C.1.pep
US-10-437-963-152994
```

```
Query Match          61.8%; Score 55; DB 16; Length 248;
Best Local Similarity 75.0%; Pred. No. 47;
Matches      9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
```

```
Qy      4 PPQPLPPTVMFP 15
        |||| |||:| |
Db      134 PPQPPPPTIMAP 145
```

#### RESULT 5

US-09-794-346-1

```
; Sequence 1, Application US/09794346
; Patent No. US20010031857A1
; GENERAL INFORMATION:
; APPLICANT: Aventis Pharma Deutschland GmbH
; TITLE OF INVENTION: Memno Peptides, Process for Their Preparation and Use
Thereof
; FILE REFERENCE: 02481.1728
; CURRENT APPLICATION NUMBER: US/09/794,346
; CURRENT FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: EP 00104114.4
; PRIOR FILING DATE: 2000-02-29
; PRIOR APPLICATION NUMBER: PCT/EP 01/01661
; PRIOR FILING DATE: 2001-02-15
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 10
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Description of Artificial Sequence: Memnoniella
echinata, FH 227
; OTHER INFORMATION: 1, DSM 1319
US-09-794-346-1
```

Query Match 59.6%; Score 53; DB 9; Length 10;  
Best Local Similarity 90.0%; Pred. No. 4.5;  
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MHQPPQPLPP 10  
|||||  
Db 1 MHQPHQPLPP 10

RESULT 6

US-10-380-371-1

; Sequence 1, Application US/10380371  
; Publication No. US20040086965A1  
; GENERAL INFORMATION:  
; APPLICANT: CLAUS, JUAN  
; APPLICANT: COMINI, MARCELO  
; APPLICANT: TONARELLI, GEORGINA  
; APPLICANT: PERIN, JUAN CARLOS  
; APPLICANT: SALVETTI, JORGE LUIS  
; APPLICANT: FRANK, RONALD  
; TITLE OF INVENTION: CASEIN PEPTIDE FRAGMENTS HAVING GROWTH-INFLUENCING  
; TITLE OF INVENTION: ACTIVITY ON CELL CULTURES  
; FILE REFERENCE: 930008-2096  
; CURRENT APPLICATION NUMBER: US/10/380,371  
; CURRENT FILING DATE: 2003-10-22  
; PRIOR APPLICATION NUMBER: PCT/DE01/03849  
; PRIOR FILING DATE: 2001-10-09  
; PRIOR APPLICATION NUMBER: DE 10050091.9  
; PRIOR FILING DATE: 2000-10-09  
; NUMBER OF SEQ ID NOS: 1  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 10  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
; OTHER INFORMATION: peptide  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)  
; OTHER INFORMATION: N-term may be acetylated  
US-10-380-371-1

Query Match 59.6%; Score 53; DB 16; Length 10;  
Best Local Similarity 90.0%; Pred. No. 4.5;  
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 HQPPQPLPPT 11  
||| |||||  
Db 1 HQPHQPLPPT 10

RESULT 7

US-09-867-550-978

; Sequence 978, Application US/09867550  
; Patent No. US20020082206A1

```
; GENERAL INFORMATION:
; APPLICANT: Leach, Martin D.
; APPLICANT: Mehraban, Fuad,
; APPLICANT: Conley, Pamela
; APPLICANT: Law, Debbie
; APPLICANT: Topper, James
; TITLE OF INVENTION: No. US20020082206A1el Polynucleotides from Atherogenic
Cells and Polypeptides Encoded
; TITLE OF INVENTION: Thereby
; FILE REFERENCE: 21402-013 (Cura-313)
; CURRENT APPLICATION NUMBER: US/09/867,550
; CURRENT FILING DATE: 2001-09-20
; PRIOR APPLICATION NUMBER: USSN 60/208,427
; PRIOR FILING DATE: 2000-05-30
; NUMBER OF SEQ ID NOS: 2125
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 978
; LENGTH: 60
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-867-550-978
```

```
Query Match          59.6%; Score 53; DB 9; Length 60;
Best Local Similarity 88.9%; Pred. No. 23;
Matches      8; Conservative 1; Mismatches      0; Indels      0; Gaps      0;
```

```
Qy          2 HQPPQPLPP 10
              |:|||||||
Db          38 HRPPQPLPP 46
```

```
RESULT 8
US-10-437-963-185370
; Sequence 185370, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules
Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 185370
; LENGTH: 2683
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
```

; LOCATION: (1)..(2683)  
; OTHER INFORMATION: unsure at all Xaa locations  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_82272C.1.pep  
US-10-437-963-185370

Query Match 59.6%; Score 53; DB 16; Length 2683;  
Best Local Similarity 64.3%; Pred. No. 6.8e+02;  
Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 2 HQPPQPLPPTVMFP 15  
|||| || ||: |  
Db 2646 HQPPVPLHPTIPXP 2659

RESULT 9

US-10-051-874-56

; Sequence 56, Application US/10051874  
; Publication No. US20040005557A1  
; GENERAL INFORMATION:

; APPLICANT: Padigaru, Muralidhara  
; APPLICANT: Alsobrook II, John P  
; APPLICANT: Colman, Steven D  
; APPLICANT: Spytek, Kimberly A  
; APPLICANT: Boldog, Ferenc  
; APPLICANT: Vernet, Corine AM  
; APPLICANT: Li, Li  
; APPLICANT: Shenoy, Suresh G  
; APPLICANT: Casman, Stacie J  
; APPLICANT: Guo, Xiaojia Sasha  
; APPLICANT: Edinger, Shlomit R  
; APPLICANT: MacDougall, John R  
; APPLICANT: Malyankar, Uriel M  
; APPLICANT: Patturajan, Meera  
; APPLICANT: Shimkets, Richard A  
; APPLICANT: Pena, Carol EA  
; APPLICANT: Tchernev, Velizar T  
; APPLICANT: Zerhusen, Bryan D  
; APPLICANT: Millet, Isabelle  
; APPLICANT: Miller, Charles E  
; APPLICANT: Lepley, Denise M  
; APPLICANT: Smithson, Glennnda  
; APPLICANT: Baumgartner, Jason C  
; APPLICANT: Herrman, John L  
; APPLICANT: Peyman, John A  
; APPLICANT: Gorman, Linda  
; APPLICANT: Mezes, Peter D  
; APPLICANT: Kekuda, Ramesh  
; APPLICANT: Taupier Jr, Raymond J  
; APPLICANT: Gerlach, Valerie  
; APPLICANT: Grosse, William M  
; APPLICANT: Liu, Xiaohong  
; APPLICANT: Ellerman, Karen  
; APPLICANT: Rothenberg, Mark  
; APPLICANT: Stone, David J  
; APPLICANT: Burgess, Catherine E

; TITLE OF INVENTION: PROTEINS, POLYNUCLEOTIDES ENCODING THEM AND METHODS OF

```

; TITLE OF INVENTION: USING THE SAME
; FILE REFERENCE: 21402-245
; CURRENT APPLICATION NUMBER: US/10/051,874
; CURRENT FILING DATE: 2002-09-25
; PRIOR APPLICATION NUMBER: 60/268,595
; PRIOR FILING DATE: 2001-02-14
; PRIOR APPLICATION NUMBER: 60/325,306
; PRIOR FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: 60/262,587
; PRIOR FILING DATE: 2001-01-18
; PRIOR APPLICATION NUMBER: 60/272,409
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: 60/262,454
; PRIOR FILING DATE: 2001-01-18
; PRIOR APPLICATION NUMBER: 60/276,777
; PRIOR FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: 60/291,672
; PRIOR FILING DATE: 2001-05-17
; PRIOR APPLICATION NUMBER: 60/330,336
; PRIOR FILING DATE: 2001-10-18
; PRIOR APPLICATION NUMBER: 60/265,530
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/261,376
; PRIOR FILING DATE: 2001-01-16
; NUMBER OF SEQ ID NOS: 269
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 56
; LENGTH: 4952
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-051-874-56

```

```

Query Match          59.6%; Score 53; DB 15; Length 4952;
Best Local Similarity 57.1%; Pred. No. 1.2e+03;
Matches      8; Conservative      3; Mismatches      3; Indels      0; Gaps      0;

```

```

Qy      1 MHQPPQPLPPTVMF 14
        |:|:|:| || | |
Db      1886 LHKPPRPQPPEVAF 1899

```

# RESULT 10

```

US-10-051-874-166
; Sequence 166, Application US/10051874
; Publication No. US20040005557A1
; GENERAL INFORMATION:
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Alsobrook II, John P
; APPLICANT: Colman, Steven D
; APPLICANT: Spytek, Kimberly A
; APPLICANT: Boldog, Ferenc
; APPLICANT: Vernet, Corine AM
; APPLICANT: Li, Li
; APPLICANT: Shenoy, Suresh G
; APPLICANT: Casman, Stacie J
; APPLICANT: Guo, Xiaojia Sasha
; APPLICANT: Edinger, Shlomit R

```



; APPLICANT: MacDougall, John R  
; APPLICANT: Malyankar, Uriel M  
; APPLICANT: Patturajan, Meera  
; APPLICANT: Shimkets, Richard A  
; APPLICANT: Pena, Carol EA  
; APPLICANT: Tchernev, Velizar T  
; APPLICANT: Zerhusen, Bryan D  
; APPLICANT: Millet, Isabelle  
; APPLICANT: Miller, Charles E  
; APPLICANT: Lepley, Denise M  
; APPLICANT: Smithson, Glennnda  
; APPLICANT: Baumgartner, Jason C  
; APPLICANT: Herrman, John L  
; APPLICANT: Peyman, John A  
; APPLICANT: Gorman, Linda  
; APPLICANT: Mezes, Peter D  
; APPLICANT: Kekuda, Ramesh  
; APPLICANT: Taupier Jr, Raymond J  
; APPLICANT: Gerlach, Valerie  
; APPLICANT: Grosse, William M  
; APPLICANT: Liu, Xiaohong  
; APPLICANT: Ellerman, Karen  
; APPLICANT: Rothenberg, Mark  
; APPLICANT: Stone, David J  
; APPLICANT: Burgess, Catherine E  
; TITLE OF INVENTION: PROTEINS, POLYNUCLEOTIDES ENCODING THEM AND METHODS OF  
; TITLE OF INVENTION: USING THE SAME  
; FILE REFERENCE: 21402-245  
; CURRENT APPLICATION NUMBER: US/10/051,874  
; CURRENT FILING DATE: 2002-09-25  
; PRIOR APPLICATION NUMBER: 60/268,595  
; PRIOR FILING DATE: 2001-02-14  
; PRIOR APPLICATION NUMBER: 60/325,306  
; PRIOR FILING DATE: 2001-09-27  
; PRIOR APPLICATION NUMBER: 60/262,587  
; PRIOR FILING DATE: 2001-01-18  
; PRIOR APPLICATION NUMBER: 60/272,409  
; PRIOR FILING DATE: 2001-02-28  
; PRIOR APPLICATION NUMBER: 60/262,454  
; PRIOR FILING DATE: 2001-01-18  
; PRIOR APPLICATION NUMBER: 60/276,777  
; PRIOR FILING DATE: 2001-03-16  
; PRIOR APPLICATION NUMBER: 60/291,672  
; PRIOR FILING DATE: 2001-05-17  
; PRIOR APPLICATION NUMBER: 60/330,336  
; PRIOR FILING DATE: 2001-10-18  
; PRIOR APPLICATION NUMBER: 60/265,530  
; PRIOR FILING DATE: 2001-01-31  
; PRIOR APPLICATION NUMBER: 60/261,376  
; PRIOR FILING DATE: 2001-01-16  
; NUMBER OF SEQ ID NOS: 269  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 166  
; LENGTH: 5008  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-051-874-166

Query Match 59.6%; Score 53; DB 15; Length 5008;  
Best Local Similarity 57.1%; Pred. No. 1.2e+03;  
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 1 MHQPPQPLPPTVMF 14  
:|:|:| | | | |  
Db 1937 LHKPPRPQPPEVAF 1950

RESULT 11

US-10-085-198-112

; Sequence 112, Application US/10085198  
; Publication No. US20040009907A1  
; GENERAL INFORMATION:  
; APPLICANT: Alsobrook et al.  
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same  
; FILE REFERENCE: 21402-279  
; CURRENT APPLICATION NUMBER: US/10/085,198  
; CURRENT FILING DATE: 2002-02-25  
; PRIOR APPLICATION NUMBER: 60/271,646  
; PRIOR FILING DATE: 2001-02-26  
; PRIOR APPLICATION NUMBER: 60/276,401  
; PRIOR FILING DATE: 2001-03-16  
; PRIOR APPLICATION NUMBER: 60/311,981  
; PRIOR FILING DATE: 2001-08-13  
; PRIOR APPLICATION NUMBER: 60/312,858  
; PRIOR FILING DATE: 2001-08-16  
; PRIOR APPLICATION NUMBER: 60/271,840  
; PRIOR FILING DATE: 2001-02-27  
; PRIOR APPLICATION NUMBER: 60/277,324  
; PRIOR FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: 60/286,096  
; PRIOR FILING DATE: 2001-04-21  
; PRIOR APPLICATION NUMBER: 60/299,695  
; PRIOR FILING DATE: 2001-06-20  
; PRIOR APPLICATION NUMBER: 60/315,614  
; PRIOR FILING DATE: 2001-08-29  
; PRIOR APPLICATION NUMBER: 60/272,405  
; PRIOR FILING DATE: 2001-02-28  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 653  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 112  
; LENGTH: 5159  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-085-198-112

Query Match 59.6%; Score 53; DB 15; Length 5159;  
Best Local Similarity 57.1%; Pred. No. 1.2e+03;  
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 1 MHQPPQPLPPTVMF 14  
:|:|:| | | | |  
Db 1886 LHKPPRPQPPEVAF 1899

RESULT 12

US-10-051-874-165

; Sequence 165, Application US/10051874

; Publication No. US20040005557A1

; GENERAL INFORMATION:

; APPLICANT: Padigaru, Muralidhara

; APPLICANT: Alsobrook II, John P

; APPLICANT: Colman, Steven D

; APPLICANT: Spytek, Kimberly A

; APPLICANT: Boldog, Ferenc

; APPLICANT: Vernet, Corine AM

; APPLICANT: Li, Li

; APPLICANT: Shenoy, Suresh G

; APPLICANT: Casman, Stacie J

; APPLICANT: Guo, Xiaojia Sasha

; APPLICANT: Edinger, Shlomit R

; APPLICANT: MacDougall, John R

; APPLICANT: Malyankar, Uriel M

; APPLICANT: Patturajan, Meera

; APPLICANT: Shimkets, Richard A

; APPLICANT: Pena, Carol EA

; APPLICANT: Tchernev, Velizar T

; APPLICANT: Zerhusen, Bryan D

; APPLICANT: Millet, Isabelle

; APPLICANT: Miller, Charles E

; APPLICANT: Lepley, Denise M

; APPLICANT: Smithson, Glennnda

; APPLICANT: Baumgartner, Jason C

; APPLICANT: Herrman, John L

; APPLICANT: Peyman, John A

; APPLICANT: Gorman, Linda

; APPLICANT: Mezes, Peter D

; APPLICANT: Kekuda, Ramesh

; APPLICANT: Taupier Jr, Raymond J

; APPLICANT: Gerlach, Valerie

; APPLICANT: Grosse, William M

; APPLICANT: Liu, Xiaohong

; APPLICANT: Ellerman, Karen

; APPLICANT: Rothenberg, Mark

; APPLICANT: Stone, David J

; APPLICANT: Burgess, Catherine E

; TITLE OF INVENTION: PROTEINS, POLYNUCLEOTIDES ENCODING THEM AND METHODS OF

; TITLE OF INVENTION: USING THE SAME

; FILE REFERENCE: 21402-245

; CURRENT APPLICATION NUMBER: US/10/051,874

; CURRENT FILING DATE: 2002-09-25

; PRIOR APPLICATION NUMBER: 60/268,595

; PRIOR FILING DATE: 2001-02-14

; PRIOR APPLICATION NUMBER: 60/325,306

; PRIOR FILING DATE: 2001-09-27

; PRIOR APPLICATION NUMBER: 60/262,587

; PRIOR FILING DATE: 2001-01-18

; PRIOR APPLICATION NUMBER: 60/272,409

; PRIOR FILING DATE: 2001-02-28

; PRIOR APPLICATION NUMBER: 60/262,454

; PRIOR FILING DATE: 2001-01-18

```
; PRIOR APPLICATION NUMBER: 60/276,777
; PRIOR FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: 60/291,672
; PRIOR FILING DATE: 2001-05-17
; PRIOR APPLICATION NUMBER: 60/330,336
; PRIOR FILING DATE: 2001-10-18
; PRIOR APPLICATION NUMBER: 60/265,530
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/261,376
; PRIOR FILING DATE: 2001-01-16
; NUMBER OF SEQ ID NOS: 269
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 165
;   LENGTH: 5262
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-051-874-165
```

```
Query Match          59.6%;   Score 53;   DB 15;   Length 5262;
Best Local Similarity 57.1%;   Pred. No. 1.2e+03;
Matches      8;   Conservative    3;   Mismatches    3;   Indels      0;   Gaps      0;
```

```
Qy          1 MHQPPQPLPPTVMF 14
              |:|:|:| || | |
Db          2191 LHKPPRPQPPEVAF 2204
```

# RESULT 13

```
US-10-051-874-167
; Sequence 167, Application US/10051874
; Publication No. US20040005557A1
; GENERAL INFORMATION:
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Alsobrook II, John P
; APPLICANT: Colman, Steven D
; APPLICANT: Spytek, Kimberly A
; APPLICANT: Boldog, Ferenc
; APPLICANT: Vernet, Corine AM
; APPLICANT: Li, Li
; APPLICANT: Shenoy, Suresh G
; APPLICANT: Casman, Stacie J
; APPLICANT: Guo, Xiaojia Sasha
; APPLICANT: Edinger, Shlomit R
; APPLICANT: MacDougall, John R
; APPLICANT: Malyankar, Uriel M
; APPLICANT: Patturajan, Meera
; APPLICANT: Shimkets, Richard A
; APPLICANT: Pena, Carol EA
; APPLICANT: Tchernev, Velizar T
; APPLICANT: Zerhusen, Bryan D
; APPLICANT: Millet, Isabelle
; APPLICANT: Miller, Charles E
; APPLICANT: Lepley, Denise M
; APPLICANT: Smithson, Glennnda
; APPLICANT: Baumgartner, Jason C
; APPLICANT: Herrman, John L
; APPLICANT: Peyman, John A
```

```

; APPLICANT: Gorman, Linda
; APPLICANT: Mezes, Peter D
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Taupier Jr, Raymond J
; APPLICANT: Gerlach, Valerie
; APPLICANT: Grosse, William M
; APPLICANT: Liu, Xiaohong
; APPLICANT: Ellerman, Karen
; APPLICANT: Rothenberg, Mark
; APPLICANT: Stone, David J
; APPLICANT: Burgess, Catherine E
; TITLE OF INVENTION: PROTEINS, POLYNUCLEOTIDES ENCODING THEM AND METHODS OF
; TITLE OF INVENTION: USING THE SAME
; FILE REFERENCE: 21402-245
; CURRENT APPLICATION NUMBER: US/10/051,874
; CURRENT FILING DATE: 2002-09-25
; PRIOR APPLICATION NUMBER: 60/268,595
; PRIOR FILING DATE: 2001-02-14
; PRIOR APPLICATION NUMBER: 60/325,306
; PRIOR FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: 60/262,587
; PRIOR FILING DATE: 2001-01-18
; PRIOR APPLICATION NUMBER: 60/272,409
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: 60/262,454
; PRIOR FILING DATE: 2001-01-18
; PRIOR APPLICATION NUMBER: 60/276,777
; PRIOR FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: 60/291,672
; PRIOR FILING DATE: 2001-05-17
; PRIOR APPLICATION NUMBER: 60/330,336
; PRIOR FILING DATE: 2001-10-18
; PRIOR APPLICATION NUMBER: 60/265,530
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/261,376
; PRIOR FILING DATE: 2001-01-16
; NUMBER OF SEQ ID NOS: 269
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 167
; LENGTH: 5262
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-051-874-167

```

```

Query Match          59.6%; Score 53; DB 15; Length 5262;
Best Local Similarity 57.1%; Pred. No. 1.2e+03;
Matches      8; Conservative      3; Mismatches      3; Indels      0; Gaps      0;

```

```

Qy      1 MHQPPQPLPPTVMF 14
        |:|:|:| || || |
Db      2191 LHKPPRPQPPEVAF 2204

```

```

RESULT 14
US-10-437-963-162903
; Sequence 162903, Application US/10437963
; Publication No. US20040123343A1

```

```
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules
Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 162903
; LENGTH: 643
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_6194C.1.pep
US-10-437-963-162903
```

```
Query Match          58.4%; Score 52; DB 16; Length 643;
Best Local Similarity 75.0%; Pred. No. 2.5e+02;
Matches      9; Conservative 0; Mismatches      3; Indels      0; Gaps      0;
```

```
Qy      4 PPQPLPPTVMFP 15
        || | ||| |||
Db      88 PPTPPPTKMF 99
```

```
RESULT 15
US-10-437-963-161687
; Sequence 161687, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules
Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 161687
; LENGTH: 95
; TYPE: PRT
; ORGANISM: Oryza sativa
```

; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_60849C.1.pep  
US-10-437-963-161687

Query Match 57.3%; Score 51; DB 16; Length 95;  
Best Local Similarity 80.0%; Pred. No. 59;  
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 HQPPQPLPPT 11  
|:|||| |||  
Db 53 HKPPQPPPT 62

Search completed: August 24, 2004, 16:41:30  
Job time : 55.291 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 24, 2004, 15:23:00 ; Search time 46.3433 Seconds  
(without alignments)  
102.124 Million cell updates/sec

Title: US-09-641-801-34  
Perfect score: 89  
Sequence: 1 MHQPPQPLPPTVMFP 15

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_25:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_rvirus:\*  
16: sp\_bacteriap:\*  
17: sp\_archeap:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

%  
Result Query  
No. Score Match Length DB ID Description  
-----



1	83	93.3	141	6	Q27953	Q27953 balaenopter
2	81	91.0	141	6	Q27939	Q27939 alces alces
3	81	91.0	141	6	Q27938	Q27938 antilocapra
4	81	91.0	223	6	Q95L76	Q95L76 capra hircu
5	80	89.9	146	6	Q9BDG5	Q9bdg5 bos taurus
6	78	87.6	141	6	P79231	P79231 physeter ca
7	75	84.3	141	6	Q28418	Q28418 giraffa cam
8	65	73.0	141	6	Q28355	Q28355 delphinapte
9	63	70.8	72	6	Q28795	Q28795 tayassu taj
10	56	62.9	412	4	Q9H5E0	Q9h5e0 homo sapien
11	55	61.8	117	6	Q28442	Q28442 hippopotamu
12	54	60.7	141	6	Q29136	Q29136 tapirus ind
13	53	59.6	145	6	Q29151	Q29151 uncia uncia
14	53	59.6	421	4	Q8TC66	Q8tc66 homo sapien
15	52	58.4	250	6	Q9N2G8	Q9n2g8 canis famil
16	52	58.4	1596	5	Q9VWC6	Q9vwc6 drosophila
17	52	58.4	1596	5	Q7YU17	Q7yu17 drosophila
18	51	57.3	2876	10	Q9FE41	Q9fe41 oryza sativ
19	50	56.2	169	5	Q8SVY7	Q8svy7 encephalito
20	50	56.2	287	4	Q9H6N8	Q9h6n8 homo sapien
21	50	56.2	295	5	Q21781	Q21781 caenorhabdi
22	50	56.2	633	4	Q96RR9	Q96rr9 homo sapien
23	50	56.2	686	3	Q872B3	Q872b3 neurospora
24	50	56.2	1270	4	Q96JH2	Q96jh2 homo sapien
25	49	55.1	211	10	Q84PY7	Q84py7 oryza sativ
26	49	55.1	292	5	O62458	O62458 caenorhabdi
27	49	55.1	337	10	Q8GYK6	Q8gyk6 arabidopsis
28	49	55.1	346	10	O50068	O50068 arabidopsis
29	49	55.1	818	11	Q8C8U8	Q8c8u8 mus musculu
30	49	55.1	1385	13	Q7ZZ45	Q7zz45 brachydanio
31	48	53.9	113	10	Q7XSL3	Q7xsl3 oryza sativ
32	48	53.9	141	6	Q28229	Q28229 camelus dro
33	48	53.9	420	5	O17057	O17057 caenorhabdi
34	48	53.9	637	5	Q86KG7	Q86kg7 dictyosteli
35	47.5	53.4	261	5	Q9VAX5	Q9vax5 drosophila
36	47	52.8	141	6	Q29139	Q29139 tragulus na
37	47	52.8	191	16	Q9A4I6	Q9a4i6 caulobacter
38	47	52.8	214	16	Q7UYU1	Q7uyul rhodopirell
39	47	52.8	270	16	Q9RMT3	Q9rmt3 pseudomonas
40	47	52.8	331	10	Q7XUZ3	Q7xuz3 oryza sativ
41	47	52.8	349	16	Q88KN8	Q88kn8 pseudomonas
42	47	52.8	400	5	Q961D1	Q96ld1 drosophila
43	47	52.8	482	2	Q7WUV5	Q7wuv5 rickettsia
44	47	52.8	487	10	Q9FKI9	Q9fki9 arabidopsis
45	47	52.8	487	12	Q805Y1	Q805y1 simian herp

#### ALIGNMENTS

#### RESULT 1

Q27953

ID Q27953 PRELIMINARY; PRT; 141 AA.  
AC Q27953;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)

DE B-casein (Fragment).  
 OS Balaenoptera physalus (Finback whale) (Common rorqual).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Cetacea; Mysticeti;  
 OC Balaenopteridae; Balaenoptera.  
 OX NCBI\_TaxID=9770;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96364219; PubMed=8752004;  
 RA Gatesy J., Hayashi C., Cronin M.A., Arctander P.;  
 RT "Evidence from milk casein genes that cetaceans are close relatives of  
 RT hippopotamid artiodactyls.";  
 RL Mol. Biol. Evol. 13:954-963(1996).  
 DR EMBL; U53900; AAB08405.1; -.  
 DR InterPro; IPR001588; Casein.  
 DR Pfam; PF00363; caseins; 1.  
 FT NON\_TER 1 1  
 FT NON\_TER 141 141  
 SQ SEQUENCE 141 AA; 15822 MW; 7C3EDEE320034513 CRC64;

Query Match 93.3%; Score 83; DB 6; Length 141;  
 Best Local Similarity 93.3%; Pred. No. 0.00014;  
 Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MHQPPQPLPPTVMFP 15  
 |||||  
 Db 96 MHQPPQPLPPTPMFP 110

## RESULT 2

Q27939

ID Q27939 PRELIMINARY; PRT; 141 AA.  
 AC Q27939;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE B-casein (Fragment).  
 OS Alces alces (moose).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Cervoidea;  
 OC Cervidae; Odocoileinae; Alces.  
 OX NCBI\_TaxID=9852;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96364219; PubMed=8752004;  
 RA Gatesy J., Hayashi C., Cronin M.A., Arctander P.;  
 RT "Evidence from milk casein genes that cetaceans are close relatives of  
 RT hippopotamid artiodactyls.";  
 RL Mol. Biol. Evol. 13:954-963(1996).  
 DR EMBL; U53896; AAB08403.1; -.  
 DR InterPro; IPR001588; Casein.  
 DR Pfam; PF00363; caseins; 1.  
 FT NON\_TER 1 1  
 FT NON\_TER 141 141  
 SQ SEQUENCE 141 AA; 15763 MW; DC39F68595C13C72 CRC64;

Query Match 91.0%; Score 81; DB 6; Length 141;

Best Local Similarity 93.3%; Pred. No. 0.00027;  
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MHQPPQPLPPTVMFP 15  
||| |||||  
Db 96 MHQTPQPLPPTVMFP 110

#### RESULT 3

Q27938

ID Q27938 PRELIMINARY; PRT; 141 AA.  
AC Q27938;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE B-casein (Fragment).  
OS Antilocapra americana (Pronghorn).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;  
OC Antilocapridae; Antilocapra.  
OX NCBI\_TaxID=9891;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96364219; PubMed=8752004;  
RA Gatesy J., Hayashi C., Cronin M.A., Arctander P.;  
RT "Evidence from milk casein genes that cetaceans are close relatives of  
RT hippopotamid artiodactyls.";  
RL Mol. Biol. Evol. 13:954-963(1996).  
DR EMBL; U53895; AAB08402.1; -.  
DR InterPro; IPR001588; Casein.  
DR Pfam; PF00363; caseins; 1.  
FT NON\_TER 1 1  
FT NON\_TER 141 141  
SQ SEQUENCE 141 AA; 15667 MW; F1112AF1617119BB CRC64;

Query Match 91.0%; Score 81; DB 6; Length 141;  
Best Local Similarity 93.3%; Pred. No. 0.00027;  
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MHQPPQPLPPTVMFP 15  
||| |||||  
Db 96 MHQSPQPLPPTVMFP 110

#### RESULT 4

Q95L76

ID Q95L76 PRELIMINARY; PRT; 223 AA.  
AC Q95L76;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
DE Beta-casein precursor.  
GN CSN2.  
OS Capra hircus (Goat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;  
OC Bovidae; Caprinae; Capra.

OX NCBI\_TaxID=9925;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Blood;  
 RA Wang Q., Huang Z., Chen M.J., Huang S.Z., Zeng Y.T.;  
 RL Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AF409096; AAK97639.1; -.  
 DR InterPro; IPR001588; Casein.  
 DR Pfam; PF00363; caseins; 1.  
 KW Signal.  
 FT SIGNAL 1 15 POTENTIAL.  
 FT CHAIN 16 223 BETA-CASEIN.  
 SQ SEQUENCE 223 AA; 24992 MW; 35A8BE17746A01DB CRC64;

Query Match 91.0%; Score 81; DB 6; Length 223;  
 Best Local Similarity 93.3%; Pred. No. 0.00042;  
 Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MHQPPQPLPPTVMFP 15  
 ||||| |||||  
 Db 159 MHQPPQPLSPTVMFP 173

#### RESULT 5

##### Q9BDG5

ID Q9BDG5 PRELIMINARY; PRT; 146 AA.  
 AC Q9BDG5;  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)  
 DE Beta casein B (Fragment).  
 GN BCN B.  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;  
 OC Bovidae; Bovinae; Bos.  
 OX NCBI\_TaxID=9913;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Mammary gland;  
 RA Klotz A., Buchberger J., Krause I., Einspanier R.;  
 RT "Characterization of milk proteins."  
 RL Submitted (JAN-2001) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AJ296330; CAC37028.1; -.  
 DR InterPro; IPR001588; Casein.  
 DR Pfam; PF00363; caseins; 1.  
 FT NON\_TER 1 1  
 FT NON\_TER 146 146  
 SQ SEQUENCE 146 AA; 16453 MW; 48A77E25740A9891 CRC64;

Query Match 89.9%; Score 80; DB 6; Length 146;  
 Best Local Similarity 93.3%; Pred. No. 0.00039;  
 Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MHQPPQPLPPTVMFP 15  
 |||| |||||  
 Db 97 MHQPHQPLPPTVMFP 111

RESULT 6

P79231

ID P79231 PRELIMINARY; PRT; 141 AA.  
 AC P79231;  
 DT 01-MAY-1997 (TrEMBLrel. 03, Created)  
 DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE Beta casein (Fragment).  
 OS Physeter catodon (Sperm whale) (Physeter macrocephalus).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Cetacea; Odontoceti;  
 OC Physeteridae; Physeter.  
 OX NCBI\_TaxID=9755;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Gatesy J.;  
 RT "More DNA support for a Cetacea/Hippopotamidae clade: the blood  
 clotting protein gene g-fibrinogen.";  
 RL Mol. Biol. Evol. 0:0-0(1997).  
 DR EMBL; U86644; AAB47430.1; -.  
 DR InterPro; IPR001588; Casein.  
 DR Pfam; PF00363; caseins; 1.  
 FT NON\_TER 1 1  
 FT NON\_TER 141 141  
 SQ SEQUENCE 141 AA; 15867 MW; 0267BA4DD8FEB9B2 CRC64;

Query Match 87.6%; Score 78; DB 6; Length 141;  
 Best Local Similarity 86.7%; Pred. No. 0.00072;  
 Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MHQPPQPLPPTVMFP 15  
 ||||| ||||| |||  
 Db 96 MHQPPHPLPPTPMFP 110

RESULT 7

Q28418

ID Q28418 PRELIMINARY; PRT; 141 AA.  
 AC Q28418;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE B-casein (Fragment).  
 OS Giraffa camelopardalis (Giraffe).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Giraffoidea;  
 OC Giraffidae; Giraffa.  
 OX NCBI\_TaxID=9894;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96364219; PubMed=8752004;  
 RA Gatesy J., Hayashi C., Cronin M.A., Arctander P.;  
 RT "Evidence from milk casein genes that cetaceans are close relatives of  
 hippopotamid artiodactyls.";  
 RL Mol. Biol. Evol. 13:954-963(1996).

DR EMBL; U53897; AAB08412.1; -.  
DR InterPro; IPR001588; Casein.  
DR Pfam; PF00363; caseins; 1.  
FT NON\_TER 1 1  
FT NON\_TER 141 141  
SQ SEQUENCE 141 AA; 15712 MW; 546DBF082A26BD73 CRC64;

Query Match 84.3%; Score 75; DB 6; Length 141;  
Best Local Similarity 86.7%; Pred. No. 0.002;  
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MHQPPQPLPPTVMFP 15  
||| |||||  
Db 96 MHQSPQPLPPTVMLP 110

RESULT 8

Q28355

ID Q28355 PRELIMINARY; PRT; 141 AA.  
AC Q28355;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE B-casein (Fragment).  
OS Delphinapterus leucas (Beluga whale).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Cetacea; Odontoceti;  
OC Monodontidae; Delphinapterus.  
OX NCBI\_TaxID=9749;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96364219; PubMed=8752004;  
RA Gatesy J., Hayashi C., Cronin M.A., Arctander P.;  
RT "Evidence from milk casein genes that cetaceans are close relatives of  
RT hippopotamid artiodactyls.";  
RL Mol. Biol. Evol. 13:954-963(1996).  
DR EMBL; U53899; AAB08408.1; -.  
DR InterPro; IPR001588; Casein.  
DR Pfam; PF00363; caseins; 1.  
FT NON\_TER 1 1  
FT NON\_TER 141 141  
SQ SEQUENCE 141 AA; 16113 MW; F2619AEE88864D5A CRC64;

Query Match 73.0%; Score 65; DB 6; Length 141;  
Best Local Similarity 73.3%; Pred. No. 0.053;  
Matches 11; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 MHQPPQPLPPTVMFP 15  
||||| |||  
Db 96 MHQPPHRFPPTPMFP 110

RESULT 9

Q28795

ID Q28795 PRELIMINARY; PRT; 72 AA.  
AC Q28795;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)

DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE B-casein (Fragment).  
 OS *Tayassu tajacu* (Collared peccary) (*Pecari tajacu*).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Tayassuidae; Pecari.  
 OX NCBI\_TaxID=9829;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96364219; PubMed=8752004;  
 RA Gatesy J., Hayashi C., Cronin M.A., Arctander P.;  
 RT "Evidence from milk casein genes that cetaceans are close relatives of  
 RT hippopotamid artiodactyls.";  
 RL Mol. Biol. Evol. 13:954-963(1996).  
 DR EMBL; U53903; AAB08417.1; -.  
 DR InterPro; IPR001588; Casein.  
 DR Pfam; PF00363; caseins; 1.  
 FT NON\_TER 1 1  
 FT NON\_TER 72 72  
 SQ SEQUENCE 72 AA; 8039 MW; 0EAB068949A1112E CRC64;

Query Match 70.8%; Score 63; DB 6; Length 72;  
 Best Local Similarity 73.3%; Pred. No. 0.054;  
 Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 MHQPPQPLPPTVMFP 15  
 ||| |||:| | |||  
 Db 51 MHQVPQPIRTPMFP 65

# RESULT 10

Q9H5E0

ID Q9H5E0 PRELIMINARY; PRT; 412 AA.  
 AC Q9H5E0;  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Hypothetical protein FLJ23531.  
 OS *Homo sapiens* (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; *Homo*.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Lung;  
 RA Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T., Matsumura K.,  
 RA Nakajima Y., Mizuno T., Morinaga M., Tanigami A., Fujiwara T., Ono T.,  
 RA Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y., Ota T., Suzuki Y.,  
 RA Obayashi M., Nishi T., Shibahara T., Tanaka T., Nakamura Y.,  
 RA Isogai T., Sugano S.;  
 RT "NEDO human cDNA sequencing project.";  
 RL Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AK027184; BAB15686.1; -.  
 KW Hypothetical protein.  
 SQ SEQUENCE 412 AA; 46539 MW; D72A6D830BB12B94 CRC64;

Query Match 62.9%; Score 56; DB 4; Length 412;

Best Local Similarity 69.2%; Pred. No. 2.9;  
Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 3 QPPQPLPPTVMFP 15  
||||| || :||  
Db 15 QPPQPAPPPPLFP 27

RESULT 11

Q28442

ID Q28442 PRELIMINARY; PRT; 117 AA.  
AC Q28442;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE B-casein (Fragment).  
OS Hippopotamus amphibius (Hippopotamus).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Hippopotamidae; Hippopotamus.  
OX NCBI\_TaxID=9833;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96364219; PubMed=8752004;  
RA Gatesy J., Hayashi C., Cronin M.A., Arctander P.;  
RT "Evidence from milk casein genes that cetaceans are close relatives of  
RT hippopotamid artiodactyls.";  
RL Mol. Biol. Evol. 13:954-963(1996).  
DR EMBL; U53901; AAB08414.1; -.  
DR InterPro; IPR001588; Casein.  
DR Pfam; PF00363; caseins; 1.  
FT NON\_TER 1 1  
FT NON\_TER 117 117  
SQ SEQUENCE 117 AA; 13179 MW; A50C6D9F6A126FC4 CRC64;

Query Match 61.8%; Score 55; DB 6; Length 117;  
Best Local Similarity 66.7%; Pred. No. 1.2;  
Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 1 MHQPPQPLPPTVMFP 15  
|| ||| ||:|||  
Db 68 MHPXSQPLSPTLMFP 82

RESULT 12

Q29136

ID Q29136 PRELIMINARY; PRT; 141 AA.  
AC Q29136;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE B-casein (Fragment).  
OS Tapirus indicus (Asiatic tapir) (Malayan tapir).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Perissodactyla; Tapiridae; Tapirus.  
OX NCBI\_TaxID=9802;  
RN [1]  
RP SEQUENCE FROM N.A.



RX MEDLINE=96364219; PubMed=8752004;  
 RA Gatesy J., Hayashi C., Cronin M.A., Arctander P.;  
 RT "Evidence from milk casein genes that cetaceans are close relatives of  
 RT hippopotamid artiodactyls.";  
 RL Mol. Biol. Evol. 13:954-963(1996).  
 DR EMBL; U53904; AAB08419.1; -.  
 DR InterPro; IPR001588; Casein.  
 DR Pfam; PF00363; caseins; 1.  
 FT NON\_TER 1 1  
 FT NON\_TER 141 141  
 SQ SEQUENCE 141 AA; 15952 MW; 2F6ECAA9B3123B4B CRC64;

Query Match 60.7%; Score 54; DB 6; Length 141;  
 Best Local Similarity 66.7%; Pred. No. 2;  
 Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 MHQPPQPLPPTVMFP 15  
 ||| |||| | : | |  
 Db 96 MHQVPQPLHQTLMLP 110

# RESULT 13

Q29151

ID Q29151 PRELIMINARY; PRT; 145 AA.  
 AC Q29151;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE B-casein (Fragment).  
 OS Uncia uncia (Snow leopard) (Panthera uncia).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Uncia.  
 OX NCBI\_TaxID=29064;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96364219; PubMed=8752004;  
 RA Gatesy J., Hayashi C., Cronin M.A., Arctander P.;  
 RT "Evidence from milk casein genes that cetaceans are close relatives of  
 RT hippopotamid artiodactyls.";  
 RL Mol. Biol. Evol. 13:954-963(1996).  
 DR EMBL; U53906; AAB08422.1; -.  
 DR InterPro; IPR001588; Casein.  
 DR Pfam; PF00363; caseins; 1.  
 FT NON\_TER 1 1  
 FT NON\_TER 145 145  
 SQ SEQUENCE 145 AA; 16379 MW; CAD027AAEE62012A CRC64;

Query Match 59.6%; Score 53; DB 6; Length 145;  
 Best Local Similarity 66.7%; Pred. No. 2.9;  
 Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 MHQPPQPLPPTVMFP 15  
 ||| || || | | |  
 Db 100 MHQIPQALPQTTMLP 114

# RESULT 14

Q8TC66

ID Q8TC66 PRELIMINARY; PRT; 421 AA.  
AC Q8TC66;  
DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Similar to hypothetical protein DKFZp547H236.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Testis;  
RA Strausberg R.;  
RL Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.  
DR EMBL; BC025404; AAH25404.1; -.  
DR GO; GO:0005634; C:nucleus; IEA.  
DR GO; GO:0003700; F:transcription factor activity; IEA.  
DR GO; GO:0005215; F:transporter activity; IEA.  
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.  
DR GO; GO:0006810; P:transport; IEA.  
DR InterPro; IPR001356; Homeobox.  
DR InterPro; IPR000566; Lipocln\_cytFABP.  
DR Pfam; PF00046; homeobox; 1.  
DR ProDom; PD000010; Homeobox; 1.  
DR SMART; SM00389; HOX; 1.  
DR PROSITE; PS50071; HOMEBOX\_2; 1.  
DR PROSITE; PS00213; LIPOCALIN; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 421 AA; 46174 MW; 3622AAB62D32561F CRC64;

Query Match 59.6%; Score 53; DB 4; Length 421;  
Best Local Similarity 88.9%; Pred. No. 8;  
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 HQPPQPLPP 10  
|:|||||||  
Db 38 HRPPQPLPP 46

RESULT 15

Q9N2G8

ID Q9N2G8 PRELIMINARY; PRT; 250 AA.  
AC Q9N2G8;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Beta-casein.  
OS Canis familiaris (Dog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
OX NCBI\_TaxID=9615;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20541290; PubMed=11092743;  
RA Watanabe M., Sugano S., Togashi T., Imai J., Uchida K., Yamaguchi R.,

RA Tateyama S.;  
RT "Molecular cloning and phylogenetic analysis of canine beta-casein."  
RL DNA Seq. 11:295-300(2000).  
DR EMBL; AB035080; BAA95931.1; -.  
DR InterPro; IPR001588; Casein.  
DR Pfam; PF00363; caseins; 1.  
DR PROSITE; PS00306; CASEIN\_ALPHA\_BETA; 1.  
SQ SEQUENCE 250 AA; 28401 MW; 1D58391E7BF97ED8 CRC64;

Query Match 58.4%; Score 52; DB 6; Length 250;  
Best Local Similarity 76.9%; Pred. No. 6.8;  
Matches 10; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MHQPPQPLPPTVM 13  
||| ||||| | |  
Db 178 MHQIPQPLPQTPM 190

Search completed: August 24, 2004, 15:51:01  
Job time : 49.3433 secs

OM protein - protein search, using sw model

Run on: August 24, 2004, 14:57:04 ; Search time 8.0597 Seconds  
 (without alignments)  
 96.908 Million cell updates/sec

Title: US-09-641-801-34  
 Perfect score: 89  
 Sequence: 1 MHQPPQPLPPTVMFP 15

Scoring table: BLOSUM62  
 Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0  
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
 Maximum Match 100%  
 Listing first 45 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a  
 score greater than or equal to the score of the result being printed,  
 and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	89	100.0	222	1	CASB_SHEEP	P11839 ovis aries
2	89	100.0	224	1	CASB_BUBBU	Q9tsi0 bubalus bub
3	81	91.0	222	1	CASB_CAPHI	P33048 capra hircu
4	80	89.9	224	1	CASB_BOVIN	P02666 bos taurus
5	55	61.8	232	1	CASB_PIG	P39037 sus scrofa
6	53	59.6	382	1	MEI3_HUMAN	Q99687 homo sapien
7	53	59.6	5262	1	MLL2_HUMAN	O14686 homo sapien
8	50	56.2	232	1	CASB_CAMDR	Q9tvd0 camelus dro
9	49	55.1	1386	1	ZAP3_MOUSE	Q9r0i7 mus musculu
10	49	55.1	1822	1	ZAP3_HUMAN	P49750 homo sapien
11	47	52.8	226	1	CASB_HUMAN	P05814 homo sapien
12	47	52.8	429	1	KNIR_DROME	P10734 drosophila
13	47	52.8	479	1	ATIN_HSV1F	P04486 herpes simp
14	47	52.8	490	1	ATIN_HSV11	P06492 herpes simp
15	47	52.8	490	1	ATIN_HSV2H	P23990 herpes simp
16	46	51.7	417	1	IRX5_HUMAN	P78411 homo sapien
17	46	51.7	613	1	GCM2_DROME	Q9vla2 drosophila

18	46	51.7	690	1	BPL1_YEAST	P48445	s biotin--p
19	46	51.7	915	1	PDB2_ARATH	O23078	arabidopsis
20	46	51.7	1094	1	S24C_HUMAN	P53992	homo sapien
21	45.5	51.1	103	1	SMS2_RANRI	P87385	rana ridibu
22	45	50.6	351	1	KLF2_RAT	Q9et58	rattus norv
23	45	50.6	354	1	KLF2_MOUSE	Q60843	mus musculu
24	45	50.6	416	1	HXD3_HUMAN	P31249	homo sapien
25	45	50.6	428	1	FXB2_MOUSE	Q64733	mus musculu
26	45	50.6	431	1	HKL3_ARATH	P48000	arabidopsis
27	45	50.6	1274	1	ENAM_MOUSE	O55196	mus musculu
28	45	50.6	2414	1	P300_HUMAN	Q09472	homo sapien
29	44.5	50.0	503	1	ANXB_MOUSE	P97384	mus musculu
30	44.5	50.0	505	1	ANXB_HUMAN	P50995	homo sapien
31	44	49.4	292	1	YQI4_CAEEL	Q09505	caenorhabdi
32	44	49.4	294	1	YQI9_CAEEL	Q09507	caenorhabdi
33	44	49.4	412	1	ALF_PETHY	O22621	petunia hyb
34	44	49.4	416	1	NH67_CAEEL	Q9xvv3	caenorhabdi
35	44	49.4	417	1	Y943_METJA	Q58353	methanococc
36	44	49.4	459	1	CF46_HUMAN	Q9y330	homo sapien
37	44	49.4	629	1	RA21_XENLA	O93310	xenopus lae
38	44	49.4	736	1	DVL2_HUMAN	O14641	homo sapien
39	44	49.4	1556	1	PROS_DROVI	Q9u6a1	drosophila
40	43.5	48.9	231	1	ASC1_MOUSE	Q02067	mus musculu
41	43.5	48.9	233	1	ASC1_RAT	P19359	rattus norv
42	43.5	48.9	510	1	DHAF_VIBHA	Q56694	vibrio harv
43	43.5	48.9	1294	1	RRPO_WCMVM	P09498	white clove
44	43	48.3	178	1	Y561_CHLPN	Q9z7z2	chlamydia p
45	43	48.3	309	1	NO75_SOYBN	P08297	glycine max

# ALIGNMENTS

## RESULT 1

### CASB\_SHEEP

ID CASB\_SHEEP STANDARD; PRT; 222 AA.  
AC P11839;  
DT 01-OCT-1989 (Rel. 12, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Beta casein precursor.  
GN CSN2.  
OS Ovis aries (Sheep).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;  
OC Bovidae; Caprinae; Ovis.  
OX NCBI\_TaxID=9940;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=89375530; PubMed=2505862;  
RA Provot C., Persuy M.A., Mercier J.-C.;  
RT "Complete nucleotide sequence of ovine beta-casein cDNA:  
RT inter-species comparison.";  
RL Biochimie 71:827-832(1989).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=95197013; PubMed=7890174;

RA Provot C., Persuy M.A., Mercier J.-C.;  
 RT "Complete sequence of the ovine beta-casein-encoding gene and  
 RT interspecies comparison.";  
 RL Gene 154:259-263(1995).  
 RN [3]  
 RP SEQUENCE OF 16-222.  
 RX MEDLINE=80046695; PubMed=499202;  
 RA Richardson B.C., Mercier J.-C.;  
 RT "The primary structure of the ovine beta-caseins.";  
 RL Eur. J. Biochem. 99:285-297(1979).  
 CC -!- FUNCTION: Important role in determination of the surface  
 CC properties of the casein micelles.  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- TISSUE SPECIFICITY: Mammary gland specific. Secreted in milk.  
 CC -!- SIMILARITY: Belongs to the beta-casein family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X16482; CAA34502.1; -.  
 DR EMBL; X79703; CAA56139.1; -.  
 DR PIR; A32979; A32979.  
 DR InterPro; IPR001588; Casein.  
 DR Pfam; PF00363; caseins; 1.  
 DR PROSITE; PS00306; CASEIN\_ALPHA\_BETA; 1.  
 KW Milk; Phosphorylation; Glycoprotein; Signal.  
 FT SIGNAL 1 15  
 FT CHAIN 16 222 BETA CASEIN.  
 FT MOD\_RES 30 30 PHOSPHORYLATION (POTENTIAL).  
 FT MOD\_RES 32 32 PHOSPHORYLATION (POTENTIAL).  
 FT MOD\_RES 33 33 PHOSPHORYLATION (POTENTIAL).  
 FT MOD\_RES 34 34 PHOSPHORYLATION (POTENTIAL).  
 FT CONFLICT 70 70 A -> T (IN REF. 3).  
 FT CONFLICT 82 82 P -> A (IN REF. 3).  
 SQ SEQUENCE 222 AA; 24875 MW; 061B4424DCB49EB1 CRC64;

Query Match 100.0%; Score 89; DB 1; Length 222;  
 Best Local Similarity 100.0%; Pred. No. 6.5e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MHQPPQPLPPTVMFP 15  
 |||||  
 Db 159 MHQPPQPLPPTVMFP 173

RESULT 2  
 CASB\_BUBBU  
 ID CASB\_BUBBU STANDARD; PRT; 224 AA.  
 AC Q9TSI0; O62824;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Beta casein precursor.  
 GN CSN2.  
 OS Bubalus bubalis (Domestic water buffalo).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;  
 OC Bovidae; Bovinae; Bubalus.  
 OX NCBI\_TaxID=89462;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Mammary gland;  
 RA Klotz A., Krause I., Einspanier R.;  
 RT "Isolation of mRNA from buffalo milk.";  
 RL Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Mammary gland;  
 RA Das P., Garg L.C.;  
 RT "cDNA cloning and sequencing of beta-casein gene in B. bubalis.";  
 RL Submitted (OCT-1998) to the EMBL/GenBank/DDBJ databases.  
 CC -!- FUNCTION: Important role in determination of the surface  
 CC properties of the casein micelles (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- TISSUE SPECIFICITY: Mammary gland specific. Secreted in milk.  
 CC -!- SIMILARITY: Belongs to the beta-casein family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AJ005165; CAA06408.1; -.  
 DR EMBL; AJ005432; CAA06535.1; -.  
 DR InterPro; IPR001588; Casein.  
 DR Pfam; PF00363; caseins; 1.  
 DR PROSITE; PS00306; CASEIN\_ALPHA\_BETA; 1.  
 KW Milk; Phosphorylation; Glycoprotein; Signal.  
 FT SIGNAL 1 15 BY SIMILARITY.  
 FT CHAIN 16 224 BETA CASEIN.  
 FT MOD\_RES 30 30 PHOSPHORYLATION (BY SIMILARITY).  
 FT MOD\_RES 32 32 PHOSPHORYLATION (BY SIMILARITY).  
 FT MOD\_RES 33 33 PHOSPHORYLATION (BY SIMILARITY).  
 FT MOD\_RES 34 34 PHOSPHORYLATION (BY SIMILARITY).  
 FT MOD\_RES 50 50 PHOSPHORYLATION (BY SIMILARITY).  
 FT CARBOHYD 70 70 O-LINKED (GALNAC. . .) (BY SIMILARITY).  
 FT CARBOHYD 72 72 O-LINKED (GALNAC. . .) (BY SIMILARITY).  
 FT CARBOHYD 95 95 O-LINKED (GALNAC. . .) (BY SIMILARITY).  
 FT CARBOHYD 183 183 O-LINKED (GALNAC. . .) (BY SIMILARITY).  
 FT CONFLICT 117 117 M -> T (IN REF. 2).  
 SQ SEQUENCE 224 AA; 25106 MW; 14FD3687DD17C5A9 CRC64;

Query Match 100.0%; Score 89; DB 1; Length 224;  
 Best Local Similarity 100.0%; Pred. No. 6.5e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHQPPQPLPPTVMFP 15  
 |||||  
 Db 159 MHQPPQPLPPTVMFP 173

RESULT 3

CASB\_CAPHI

ID CASB\_CAPHI STANDARD; PRT; 222 AA.

AC P33048;

DT 01-OCT-1993 (Rel. 27, Created)

DT 01-OCT-1993 (Rel. 27, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Beta casein precursor.

GN CSN2.

OS Capra hircus (Goat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;

OC Bovidae; Caprinae; Capra.

OX NCBI\_TaxID=9925;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Saanen; TISSUE=Blood;

RX MEDLINE=93077039; PubMed=1446822;

RA Roberts B., Ditullio P., Vitale J., Hehir K., Gordon K.;

RT "Cloning of the goat beta-casein-encoding gene and expression in  
 transgenic mice.";

RL Gene 121:255-262(1992).

CC -!- FUNCTION: Important role in determination of the surface  
 properties of the casein micelles.

CC -!- SUBCELLULAR LOCATION: Secreted.

CC -!- TISSUE SPECIFICITY: Mammary gland specific. Secreted in milk.

CC -!- SIMILARITY: Belongs to the beta-casein family.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; M90561; AAA30906.1; -.

DR EMBL; M90556; AAA30906.1; JOINED.

DR EMBL; M90557; AAA30906.1; JOINED.

DR EMBL; M90558; AAA30906.1; JOINED.

DR EMBL; M90560; AAA30906.1; JOINED.

DR PIR; JC1384; JC1384.

DR InterPro; IPR001588; Casein.

DR Pfam; PF00363; caseins; 1.

DR PROSITE; PS00306; CASEIN\_ALPHA\_BETA; FALSE NEG.

KW Milk; Phosphorylation; Glycoprotein; Signal.

FT SIGNAL 1 15

FT CHAIN 16 222 BETA CASEIN.

FT MOD\_RES 30 30 PHOSPHORYLATION (BY SIMILARITY).

FT MOD\_RES 32 32 PHOSPHORYLATION (BY SIMILARITY).

FT MOD\_RES 33 33 PHOSPHORYLATION (BY SIMILARITY).

FT MOD\_RES 34 34 PHOSPHORYLATION (BY SIMILARITY).



FT MOD\_RES 50 50 PHOSPHORYLATION (BY SIMILARITY).  
SQ SEQUENCE 222 AA; 24865 MW; 96AE17746A01CD05 CRC64;

Query Match 91.0%; Score 81; DB 1; Length 222;  
Best Local Similarity 93.3%; Pred. No. 0.00069;  
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MHQPPQPLPPTVMFP 15  
| | | | | | | | | | | | | | | |  
Db 159 MHQPPQPLSPTVMFP 173

#### RESULT 4

##### CASB\_BOVIN

ID CASB\_BOVIN STANDARD; PRT; 224 AA.  
AC P02666;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 01-MAR-1989 (Rel. 10, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Beta casein precursor.  
GN CSN2.  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;  
OC Bovidae; Bovinae; Bos.  
OX NCBI\_TaxID=9913;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Baev A.A., Smirnov I.K., Gorodetsky S.I.;  
RT "Primary structure of bovine beta-casein cDNA.";  
RL Mol. Biol. (Mosk) 21:214-222(1987).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=88188989; PubMed=2833669;  
RA Stewart A.F., Bonsing J., Beattie C.W., Shah F., Willis I.M.,  
RA Mackinlay A.G.;  
RT "Complete nucleotide sequences of bovine alpha S2- and beta-casein  
RT cDNAs: comparisons with related sequences in other species.";  
RL Mol. Biol. Evol. 4:231-241(1987).  
RN [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=90147279; PubMed=3271384;  
RA Bonsing J., Ring J.M., Stewart A.F., Mackinlay A.G.;  
RT "Complete nucleotide sequence of the bovine beta-casein gene.";  
RL Aust. J. Biol. Sci. 41:527-537(1988).  
RN [4]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=87128158; PubMed=3814153;  
RA Jimenez-Flores R., Kang Y.C., Richardson T.;  
RT "Cloning and sequence analysis of bovine beta-casein cDNA.";  
RL Biochem. Biophys. Res. Commun. 142:617-621(1987).  
RN [5]  
RP SEQUENCE FROM N.A. (VARIANT A3).  
RC TISSUE=Mammary gland;  
RX MEDLINE=94068382; PubMed=8248100;  
RA Simons G., van den Heuvel W., Reynen T., Frijters A., Rutten G.,  
RA Slangen C.J., Groenen M., de Vos W.M., Siezen R.J.;

RT "Overproduction of bovine beta-casein in Escherichia coli and  
 RT engineering of its main chymosin cleavage site.";  
 RL Protein Eng. 6:763-770(1993).  
 RN [6]  
 RP SEQUENCE OF 16-224 (VARIANT A2).  
 RX MEDLINE=88152252; PubMed=3278933;  
 RA Carles C., Huet J.-C., Ribadeau-Dumas B.;  
 RT "A new strategy for primary structure determination of proteins:  
 RT application to bovine beta-casein.";  
 RL FEBS Lett. 229:265-272(1988).  
 RN [7]  
 RP SEQUENCE OF 16-224 (VARIANT A2).  
 RX MEDLINE=72233212; PubMed=4557764;  
 RA Ribadeau-Dumas B., Brignon G., Grosclaude F., Mercier J.-C.;  
 RT "Primary structure of bovine beta casein. Complete sequence.";  
 RL Eur. J. Biochem. 25:505-514(1972).  
 RN [8]  
 RP VARIANTS A1; B AND C.  
 RX MEDLINE=72214259; PubMed=5064450;  
 RA Grosclaude F., Mahe M.-F., Mercier J.-C., Ribadeau-Dumas B.;  
 RT "Characterization of genetic variants of alpha-S1 and beta bovine  
 RT caseins.";  
 RL Eur. J. Biochem. 26:328-337(1972).  
 RN [9]  
 RP SEQUENCE OF 118-124 (VARIANT A3).  
 RX MEDLINE=71252171; PubMed=4997616;  
 RA Ribadeau-Dumas B., Grosclaude F., Mercier J.-C.;  
 RT "Localization in the peptide chain of bovine beta casein of the  
 RT His-Gln substitution differentiating the A2 and A3 genetic  
 RT variants.";  
 RL C. R. Acad. Sci., D, Sci. Nat. 270:2369-2372(1970).  
 RN [10]  
 RP SEQUENCE OF 48-63 (VARIANT E).  
 RX MEDLINE=75005247; PubMed=4411121;  
 RA Grosclaude F., Mahe M.-F., Voglino G.-F.;  
 RT "The beta E variant and the phosphorylation code of bovine caseins.";  
 RL FEBS Lett. 45:3-5(1974).  
 RN [11]  
 RP SEQUENCE OF 68-105 FROM N.A.  
 RX MEDLINE=85155504; PubMed=6397405;  
 RA Ivanov V.N., Kershulite D.R., Bayev A.A., Akhundova A.A.,  
 RA Sulimova G.E., Judinkova E.S., Gorodetsky S.I.;  
 RT "Identification of bacterial clones encoding bovine caseins by direct  
 RT immunological screening of the cDNA library.";  
 RL Gene 32:381-388(1984).  
 RN [12]  
 RP SEQUENCE OF 68-95 FROM N.A.  
 RX MEDLINE=86014005; PubMed=3900695;  
 RA Ivanov V.N., Kershulite D.R., Bayev A.A., Akhundova A.A.,  
 RA Silimova G.E.;  
 RT "Identification of bacterial clones coding for bovine caseins by  
 RT direct immunologic screening of the cDNA library.";  
 RL Mol. Biol. (Mosk) 19:955-963(1985).  
 RN [13]  
 RP SEQUENCE OF 18-57 FROM N.A., AND SEQUENCE OF 16-224 (VARIANT H).  
 RX MEDLINE=20154951; PubMed=10690361;  
 RA Han S.K., Shin Y.C., Byun H.D.;

RT "Biochemical, molecular and physiological characterization of a new  
 RT beta-casein variant detected in Korean cattle.";  
 RL Anim. Genet. 31:49-51(2000).  
 RN [14]  
 RP SEQUENCE OF 125-195 (VARIANTS A1 AND G).  
 RA Dong C., Ng-Kwai-Hang K.F.;  
 RT "Characterization of a non-electrophoretic genetic variant of beta-  
 RT casein by peptide mapping and mass spectrometric analysis.";  
 RL Int. Dairy J. 8:967-972(1998).  
 RN [15]  
 RP SEQUENCE OF 160-171 (VARIANT F).  
 RX MEDLINE=96118672; PubMed=7496485;  
 RA Visser S., Slangen C.J., Lagerwerf F.M., Van Dongen W.D.,  
 RA Haverkamp J.;  
 RT "Identification of a new genetic variant of bovine beta-casein using  
 RT reversed-phase high-performance liquid chromatography and mass  
 RT spectrometric analysis.";  
 RL J. Chromatogr. A 711:141-150(1995).  
 RN [16]  
 RP SEQUENCE OF 170-184 FROM N.A.  
 RX MEDLINE=83182023; PubMed=6897774;  
 RA Willis I.M., Stewart A.F., Caputo A., Thompson A.R., McKinlay A.G.;  
 RT "Construction and identification by partial nucleotide sequence  
 RT analysis of bovine casein and beta-lactoglobulin cDNA clones.";  
 RL DNA 1:375-386(1982).  
 RN [17]  
 RP CARBOHYDRATE-LINKAGE SITES.  
 RX MEDLINE=85000478; PubMed=6148101;  
 RA Yan S.B., Wold F.;  
 RT "Neoglycoproteins: in vitro introduction of glycosyl units at  
 RT glutamines in beta-casein using transglutaminase.";  
 RL Biochemistry 23:3759-3765(1984).  
 CC -!- FUNCTION: Important role in determination of the surface  
 CC properties of the casein micelles.  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- TISSUE SPECIFICITY: Mammary gland specific. Secreted in milk.  
 CC -!- POLYMORPHISM: Leu-152 is present in the variants F and G; Gln-190  
 CC and Glu-210 are present in the variant H. The sequence shown is  
 CC the A2 variant.  
 CC -!- SIMILARITY: Belongs to the beta-casein family.  
 CC -!- DATABASE: NAME=Protein Spotlight;  
 CC NOTE=Issue 16 of November 2001;  
 CC WWW="http://www.expasy.org/spotlight/articles/sptlt016.html".  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M16645; AAA30480.1; -.  
 DR EMBL; M15132; AAA30430.1; -.  
 DR EMBL; K01087; AAA30481.1; -.  
 DR EMBL; X06359; CAA29658.1; -.  
 DR EMBL; M55158; AAA30431.1; -.

DR EMBL; S67277; AAB29137.1; -.  
 DR EMBL; AF104929; AAD09813.1; -.  
 DR EMBL; AF104928; AAD09813.1; JOINED.  
 DR EMBL; M64756; AAB59254.1; -.  
 DR PIR; I45873; KBBOA2.  
 DR InterPro; IPR001588; Casein.  
 DR Pfam; PF00363; caseins; 1.  
 DR PROSITE; PS00306; CASEIN\_ALPHA\_BETA; 1.  
 KW Milk; Phosphorylation; Glycoprotein; Signal; Polymorphism.  
 FT SIGNAL 1 15  
 FT CHAIN 16 224 BETA CASEIN.  
 FT MOD\_RES 30 30 PHOSPHORYLATION.  
 FT MOD\_RES 32 32 PHOSPHORYLATION.  
 FT MOD\_RES 33 33 PHOSPHORYLATION.  
 FT MOD\_RES 34 34 PHOSPHORYLATION.  
 FT MOD\_RES 50 50 PHOSPHORYLATION (IN VARIANT A1, VARIANT  
 FT A2, VARIANT A3, VARIANT B, VARIANT E,  
 FT VARIANT F, VARIANT G AND VARIANT H).  
 FT CARBOHYD 70 70 O-LINKED (GALNAC. . .) (PARTIAL).  
 FT CARBOHYD 72 72 O-LINKED (GALNAC. . .).  
 FT CARBOHYD 95 95 O-LINKED (GALNAC. . .).  
 FT CARBOHYD 183 183 O-LINKED (GALNAC. . .).  
 FT VARIANT 40 40 R -> C (IN VARIANT H).  
 FT VARIANT 51 51 E -> K (IN VARIANT E).  
 FT VARIANT 52 52 E -> K (IN VARIANT C).  
 FT VARIANT 82 82 P -> H (IN VARIANTS A1, B, C, F AND G).  
 FT VARIANT 103 103 L -> I (IN VARIANT H).  
 FT VARIANT 121 121 H -> Q (IN VARIANT A3).  
 FT VARIANT 132 132 E -> Q (IN VARIANTS A1 AND G).  
 FT VARIANT 137 137 S -> R (IN VARIANT B).  
 FT VARIANT 152 153 LP -> PL (IN VARIANTS A1 AND H).  
 FT VARIANT 153 153 P -> L (IN VARIANT G).  
 FT VARIANT 167 167 P -> L (IN VARIANT F).  
 FT VARIANT 190 190 Q -> E (IN VARIANTS A1 AND G).  
 FT CONFLICT 108 108 M -> L (IN REF. 4 AND 7).  
 FT CONFLICT 210 210 E -> Q (IN REF. 4 AND 7).  
 FT CONFLICT 215 224 PVRGPFPIIV -> DPSLLL (IN REF. 1).  
 SQ SEQUENCE 224 AA; 25107 MW; F0BBDD8148A238AE CRC64;

Query Match 89.9%; Score 80; DB 1; Length 224;  
 Best Local Similarity 93.3%; Pred. No. 0.00093;  
 Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MHQPPQPLPPTVMFP 15  
 |||| |||||  
 Db 159 MHQPHQPLPPTVMFP 173

RESULT 5  
 CASB\_PIG  
 ID CASB\_PIG STANDARD; PRT; 232 AA.  
 AC P39037;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Beta casein precursor.  
 GN CSN2.

OS Sus scrofa (Pig).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
 OX NCBI\_TaxID=9823;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Mammary gland;  
 RX MEDLINE=92367961; PubMed=1503277;  
 RA Alexander L.J., Beattie C.W.;  
 RT "The sequence of porcine beta-casein cDNA.";  
 RL Anim. Genet. 23:369-371(1992).  
 RN [2]  
 RP SEQUENCE OF 16-29.  
 RC TISSUE=Milk;  
 RX MEDLINE=22152288; PubMed=12162653;  
 RA Kauf A.C.W., Kensinger R.S.;  
 RT "Purification of porcine beta-casein, N-terminal sequence,  
 RT quantification in mastitic milk.";  
 RL J. Anim. Sci. 80:1863-1870(2002).  
 RN [3]  
 RP CHARACTERIZATION.  
 RC TISSUE=Milk;  
 RX MEDLINE=80021173; PubMed=385058;  
 RA Mulvihill D.M., Fox P.F.;  
 RT "Isolation and characterization of porcine beta-casein.";  
 RL Biochim. Biophys. Acta 578:317-324(1979).  
 CC -!- FUNCTION: Important role in determination of the surface  
 CC properties of the casein micelles.  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- TISSUE SPECIFICITY: Mammary gland specific. Secreted in milk.  
 CC -!- SIMILARITY: Belongs to the beta-casein family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X54974; CAA38718.1; -.  
 DR InterPro; IPR001588; Casein.  
 DR Pfam; PF00363; caseins; 1.  
 DR PROSITE; PS00306; CASEIN ALPHA BETA; FALSE\_NEG.  
 KW Milk; Phosphorylation; Glycoprotein; Signal.  
 FT SIGNAL 1 15  
 FT CHAIN 16 232 BETA CASEIN.  
 FT MOD\_RES 30 30 PHOSPHORYLATION (BY SIMILARITY).  
 FT MOD\_RES 32 32 PHOSPHORYLATION (BY SIMILARITY).  
 FT MOD\_RES 33 33 PHOSPHORYLATION (BY SIMILARITY).  
 FT MOD\_RES 34 34 PHOSPHORYLATION (BY SIMILARITY).  
 FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 232 AA; 25949 MW; 6284850F40F7365C CRC64;

Query Match 61.8%; Score 55; DB 1; Length 232;  
 Best Local Similarity 71.4%; Pred. No. 1.6;  
 Matches 10; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy            1 MHQPPQPLPPTVMF 14  
              ||| |||:| | ||  
 Db           158 MHQIPQPVPQTPMF 171

RESULT 6

MEI3\_HUMAN

ID MEI3\_HUMAN            STANDARD;            PRT;    382 AA.  
 AC Q99687; Q9NPW2;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Homeobox protein Meis3 (Meis1-related protein 2) (Fragment).  
 GN MEIS3 OR MRG2.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RA Blum H., Bauersachs S., Mewes H.-W., Weil B., Wiemann S.;  
 RL Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.  
 RN [2]  
 RP SEQUENCE OF 175-382 FROM N.A.  
 RX MEDLINE=97202105; PubMed=9049632;  
 RA Steelman S., Moskow J.J., Muzynski K., North C., Druck T.,  
 RA Montgomery J.C., Huebner K., Daar I.O., Buchberg A.M.;  
 RT "Identification of a conserved family of Meis1-related homeobox  
 RT genes.";  
 RL Genome Res. 7:142-156(1997).  
 CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).  
 CC -!- SIMILARITY: Belongs to the TALE/MEIS homeobox family.  
 CC -!- SIMILARITY: Contains 1 homeobox domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AL359938; CAB95771.1; -.  
 DR EMBL; U68385; AAB19195.1; -.  
 DR HSSP; P40424; 1B72.  
 DR TRANSFAC; T03412; -.  
 DR Genew; HGNC:7002; MEIS3.  
 DR GO; GO:0005634; C:nucleus; ISS.  
 DR GO; GO:0003677; F:DNA binding; ISS.  
 DR GO; GO:0008283; P:cell proliferation; ISS.  
 DR InterPro; IPR001356; Homeobox.  
 DR Pfam; PF00046; homeobox; 1.  
 DR ProDom; PD000010; Homeobox; 1.  
 DR SMART; SM00389; HOX; 1.  
 DR PROSITE; PS00027; HOMEBOX\_1; FALSE\_NEG.

DR PROSITE; PS50071; HOMEBOX\_2; 1.  
KW DNA-binding; Nuclear protein; Homeobox.  
FT NON\_TER 1 1  
FT DOMAIN 217 256 SER/THR-RICH.  
FT DOMAIN 259 266 ASP/GLU-RICH (ACIDIC).  
FT DNA\_BIND 269 331 HOMEBOX (TALE-TYPE).  
FT CONFLICT 175 176 KM -> RP (IN REF. 2).  
FT CONFLICT 209 209 M -> I (IN REF. 2).  
FT CONFLICT 245 245 D -> V (IN REF. 2).  
FT CONFLICT 267 267 R -> P (IN REF. 2).  
FT CONFLICT 358 358 Q -> E (IN REF. 2).  
FT CONFLICT 363 367 VRPPG -> FRAPA (IN REF. 2).  
FT CONFLICT 371 377 MSLNLEG -> DEFGTRKE (IN REF. 2).  
SQ SEQUENCE 382 AA; 41821 MW; A2C11BE8061FB718 CRC64;

Query Match 59.6%; Score 53; DB 1; Length 382;  
Best Local Similarity 88.9%; Pred. No. 4.8;  
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 HQPPQPLPP 10  
|:|||||||  
Db 45 HRPPQPLPP 53

#### RESULT 7

##### MLL2\_HUMAN

ID MLL2\_HUMAN STANDARD; PRT; 5262 AA.  
AC O14686; O14687;  
DT 10-OCT-2003 (Rel. 42, Created)  
DT 10-OCT-2003 (Rel. 42, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Myeloid/lymphoid or mixed-lineage leukemia protein 2 (ALL1-related protein).  
GN MLL2 OR ALR.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3).  
RX MEDLINE=97388474; PubMed=9247308;  
RA Prasad R., Zhadanov A.B., Sedkov Y., Bullrich F., Druck T.,  
RA Rallapalli R., Yano T., Alder H., Croce C.M., Huebner K., Mazo A.,  
RA Canaani E.;  
RT "Structure and expression pattern of human ALR, a novel gene with  
RT strong homology to ALL-1 involved in acute leukemia and to Drosophila  
RT trithorax.";  
RL Oncogene 15:549-560(1997).  
RN [2]  
RP INTERACTION WITH ASC-2/NCOA6 CONTAINING COMPLEX.  
RC TISSUE=Cervical carcinoma;  
RX MEDLINE=22371496; PubMed=12482968;  
RA Goo Y.-H., Sohn Y.C., Kim D.-H., Kim S.-W., Kang M.-J., Jung D.-J.,  
RA Kwak E., Barlev N.A., Berger S.L., Chow V.T., Roeder R.G.,  
RA Azorsa D.O., Meltzer P.S., Suh P.-G., Song E.J., Lee K.-J., Lee Y.C.,  
RA Lee J.W.;  
RT "Activating signal cointegrator 2 belongs to a novel steady-state

RT complex that contains a subset of trithorax group proteins.";  
 RL Mol. Cell. Biol. 23:140-149(2003).  
 CC -!- FUNCTION: May be involved in transcriptional regulation.  
 CC -!- SUBUNIT: Belongs to the ASC-2/NCOA6 complex (ASCOM), which  
 CC contains ASC-2/NCOA6, the retinoblastoma-binding protein RBQ-3/  
 CC RBBP5, alpha- and beta-tubulins, the trithorax group proteins  
 CC MLL2 and MLL3, and ASH2/ASCL2.  
 CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).  
 CC -!- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=3;  
 CC Name=1;  
 CC IsoId=O14686-1; Sequence=Displayed;  
 CC Name=2;  
 CC IsoId=O14686-2; Sequence=VSP\_008563, VSP\_008559;  
 CC Name=3;  
 CC IsoId=O14686-3; Sequence=VSP\_008560;  
 CC -!- TISSUE SPECIFICITY: Expressed in most adult tissues, including a  
 CC variety of hematoipoietic cells, with the exception of the liver.  
 CC -!- MISCELLANEOUS: This gene mapped to a chromosomal region involved  
 CC in duplications and translocations associated with cancer.  
 CC -!- SIMILARITY: Belongs to the transcription factor trithorax family.  
 CC -!- SIMILARITY: Contains 5 PHD-type zinc fingers.  
 CC -!- SIMILARITY: Contains 1 post-SET domain.  
 CC -!- SIMILARITY: Contains 1 RING-type zinc finger.  
 CC -!- SIMILARITY: Contains 1 SET domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF010403; AAC51734.1; -.  
 DR EMBL; AF010404; AAC51735.1; -.  
 DR PIR; T03454; T03454.  
 DR PIR; T03455; T03455.  
 DR Genew; HGNC:7133; MLL2.  
 DR MIM; 602113; -.  
 DR GO; GO:0005634; C:nucleus; TAS.  
 DR GO; GO:0003700; F:transcription factor activity; TAS.  
 DR GO; GO:0007048; P:oncogenesis; TAS.  
 DR GO; GO:0006366; P:transcription from Pol II promoter; TAS.  
 DR InterPro; IPR003889; FYrich\_C.  
 DR InterPro; IPR003888; FYrich\_N.  
 DR InterPro; IPR000910; HMG\_12\_box.  
 DR InterPro; IPR003616; PostSET.  
 DR InterPro; IPR006118; Recombinase.  
 DR InterPro; IPR001214; SET.  
 DR InterPro; IPR001965; Znf\_PHD.  
 DR InterPro; IPR001841; Znf\_ring.  
 DR Pfam; PF00628; PHD; 5.  
 DR Pfam; PF00856; SET; 1.  
 DR SMART; SM00542; FYRC; 1.  
 DR SMART; SM00541; FYRN; 1.  
 DR SMART; SM00398; HMG; 1.



DR SMART; SM00249; PHD; 7.  
 DR SMART; SM00508; PostSET; 1.  
 DR SMART; SM00184; RING; 3.  
 DR SMART; SM00317; SET; 1.  
 DR PROSITE; PS50868; POST\_SET; 1.  
 DR PROSITE; PS50280; SET; 1.  
 DR PROSITE; PS01359; ZF\_PHD\_1; 5.  
 DR PROSITE; PS50016; ZF\_PHD\_2; 5.  
 DR PROSITE; PS50089; ZF\_RING\_2; 1.  
 KW Nuclear protein; Transcription regulation; Coiled coil; Zinc-finger;  
 KW Repeat; Alternative splicing; Polymorphism.  
 FT ZN\_FING 226 276 PHD-TYPE 1.  
 FT ZN\_FING 229 274 RING-TYPE.  
 FT ZN\_FING 273 323 PHD-TYPE 2.  
 FT ZN\_FING 1102 1155 PHD-TYPE 3.  
 FT ZN\_FING 1152 1202 PHD-TYPE 4.  
 FT ZN\_FING 1229 1284 PHD-TYPE 5.  
 FT DOMAIN 5121 5242 SET.  
 FT DOMAIN 5246 5262 POST-SET.  
 FT DOMAIN 2397 2436 COILED COIL (POTENTIAL).  
 FT DOMAIN 2788 2809 COILED COIL (POTENTIAL).  
 FT DOMAIN 2974 3001 COILED COIL (POTENTIAL).  
 FT DOMAIN 3286 3342 COILED COIL (POTENTIAL).  
 FT DOMAIN 3437 3476 COILED COIL (POTENTIAL).  
 FT DOMAIN 3621 3701 COILED COIL (POTENTIAL).  
 FT DOMAIN 4265 4287 COILED COIL (POTENTIAL).  
 FT DOMAIN 439 668 15 X 5 AA REPEATS OF S/P-P-P-E/P-E/A.  
 FT REPEAT 442 446 1.  
 FT REPEAT 460 464 2.  
 FT REPEAT 469 473 3.  
 FT REPEAT 496 500 4.  
 FT REPEAT 504 508 5.  
 FT REPEAT 521 525 6.  
 FT REPEAT 555 559 7.  
 FT REPEAT 564 568 8.  
 FT REPEAT 573 577 9.  
 FT REPEAT 582 586 10.  
 FT REPEAT 609 613 11.  
 FT REPEAT 618 622 12.  
 FT REPEAT 627 631 13.  
 FT REPEAT 645 649 14.  
 FT REPEAT 663 667 15.  
 FT DOMAIN 229 326 CYS-RICH.  
 FT DOMAIN 374 922 PRO-RICH.  
 FT DOMAIN 1015 1053 ARG-RICH.  
 FT DOMAIN 1122 1235 CYS-RICH.  
 FT DOMAIN 1832 2351 PRO-RICH.  
 FT DOMAIN 2536 2547 GLN-RICH.  
 FT DOMAIN 2587 2703 PRO-RICH.  
 FT DOMAIN 2986 4000 GLN-RICH.  
 FT DOMAIN 3966 4085 PRO-RICH.  
 FT DOMAIN 4634 4702 PRO-RICH.  
 FT VARSPLIC 1 305 Missing (in isoform 2).  
 FT /FTId=VSP\_008563.  
 FT VARSPLIC 306 672 PMEELPAHSWKCKACRVCACGAGSAELNPNSEWFENYSLC  
 FT HRCHKAQGGQTIRSVAEQHTPVCSRFSPPPEPGDTPTEPDPA  
 FT LYVACQGQPKGGHVTSMQPKPEGPLQCEAKPLGKAGVQLEP

FT QLEAPLNEEMPLLLPPPEESPLSPPPEESPTSPPEASRLSP  
 FT PPEELPASPLPEALHLSRPLEESPLSPPPEESPLSPPPESS  
 FT PFSPLEESPLSPPPEESPPSPALETPLSPPPEASPLSPPFEE  
 FT SPLSPPPEELPTSPPEASRLSPPPEESPMSPPEESPMSP  
 FT PPEASRLFPPEESPLSPPPEESPLSPPPEASRLSPPPEDS  
 FT PMSPPPEESPMSPPEVSRLSPLPVVSRLSPPPEESPLS  
 FT -> MSPPPEESPMSPPEASRLFPPEESPLSPPPEESPLS  
 FT PPPEASRLSPPPEDSPMSPPPEESPMSPPEVSRLSPLPVV  
 FT SRLSPPPEESPLSPPPEESPTSPPEASRLSPPPEDSPTSP  
 FT PPEDSPASPPPEDSLMSLPLEESPLLPLPEEPQLCPRSEGP  
 FT HLSRPEEPHLSRPEEPHLSPOAEHPHLSPOPEEPCLCAV  
 FT PEEPHLSPQAEHPHLSPOPEELHLSPOTEEPHLSPVPEEPC  
 FT LSPQPEESHLSPOSEEPCLSPRPEESHLSPELEKPPLSRP  
 FT EKPPEEPGQCPAPEELPLFPPPGEPSSLPLLGEPALSEPGE  
 FT PPLSPLPEELPLSPSGEPSSLSPQLMPPDPLPPPLSPIITAA  
 FT A (in isoform 2).  
 FT /FTId=VSP\_008559.  
 FT VARSPLIC 1454 1454 E -> EGET (in isoform 3).  
 FT /FTId=VSP\_008560.  
 FT VARIANT 4949 4949 R -> H (in dbSNP:3782356).  
 FT /FTId=VAR\_017115.  
 SQ SEQUENCE 5262 AA; 564171 MW; 26B7C74CAD417E44 CRC64;

Query Match 59.6%; Score 53; DB 1; Length 5262;  
 Best Local Similarity 57.1%; Pred. No. 70;  
 Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 1 MHQPPQPLPPTVMF 14  
 :|:|:| | | | |  
 Db 2191 LHKPPRPQPPEVAF 2204

#### RESULT 8

##### CASB\_CAMDR

ID CASB\_CAMDR STANDARD; PRT; 232 AA.  
 AC Q9TVD0;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Beta casein precursor.  
 GN CSN2.  
 OS Camelus dromedarius (Dromedary) (Arabian camel).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Tylopoda; Camelidae; Camelus.  
 OX NCBI\_TaxID=9838;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Somali; TISSUE=Mammary gland;  
 RX MEDLINE=98291310; PubMed=9627840;  
 RA Kappeler S., Farah Z., Puhon Z.;  
 RT "Sequence analysis of Camelus dromedarius milk caseins."  
 RL J. Dairy Res. 65:209-222(1998).  
 CC -!- FUNCTION: Important role in determination of the surface  
 CC properties of the casein micelles (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- TISSUE SPECIFICITY: Mammary gland specific. Secreted in milk.  
 CC -!- SIMILARITY: Belongs to the beta-casein family.

CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; AJ012630; CAA10079.1; -.  
DR InterPro; IPR001588; Casein.  
DR Pfam; PF00363; caseins; 1.  
DR PROSITE; PS00306; CASEIN\_ALPHA\_BETA; FALSE\_NEG.  
KW Milk; Phosphorylation; Glycoprotein; Signal.  
FT SIGNAL 1 15 BY SIMILARITY.  
FT CHAIN 16 232 BETA CASEIN.  
FT MOD\_RES 30 30 PHOSPHORYLATION (BY SIMILARITY).  
FT MOD\_RES 32 32 PHOSPHORYLATION (BY SIMILARITY).  
FT MOD\_RES 33 33 PHOSPHORYLATION (BY SIMILARITY).  
FT MOD\_RES 34 34 PHOSPHORYLATION (BY SIMILARITY).  
SQ SEQUENCE 232 AA; 26218 MW; A0F9F41D2EA7C518 CRC64;

Query Match 56.2%; Score 50; DB 1; Length 232;  
Best Local Similarity 60.0%; Pred. No. 6.9;  
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 1 MHQPPQPLPPTVMFP 15  
|:| |||:| | | |  
Db 160 MYQIPQVPQTPMIP 174

RESULT 9

ZAP3\_MOUSE

ID ZAP3\_MOUSE STANDARD; PRT; 1386 AA.  
AC Q9R0I7;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Nuclear protein ZAP3.  
GN ZAP3 OR ZAP.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Fetal liver;  
RA Misawa K., Nosaka T., Kitamura T.;  
RT "A huge nuclear protein rich in proline similar to human hypothetical  
RT protein zap3 and zap113.";  
RL Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.  
CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).  
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; AB033168; BAA85182.1; -.  
DR MGD; MGI:1926195; Zap3.  
DR GO; GO:0005634; C:nucleus; IDA.  
KW Nuclear protein.  
FT DOMAIN 15 204 PRO-RICH.  
FT DOMAIN 355 473 GLN-RICH.  
FT DOMAIN 925 1012 ARG-RICH.  
SQ SEQUENCE 1386 AA; 155130 MW; D862F9918ED221DF CRC64;

Query Match 55.1%; Score 49; DB 1; Length 1386;  
Best Local Similarity 57.1%; Pred. No. 58;  
Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 2 HQPPQPLPPTVMFP 15  
| || ||| : |  
Db 81 HLPPLPPPPVMP 94

#### RESULT 10

##### ZAP3\_HUMAN

ID ZAP3\_HUMAN STANDARD; PRT; 1822 AA.  
AC P49750; P49752; Q9P1V7;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Nuclear protein ZAP3 (ZAP113).  
GN ZAP3.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Rowen L., Madan A., Qin S., Abbasi N., Baradarani L., Birditt B.,  
RA Bloom S., Dors M., Dickhoff R., Fleetwood P., Harrison G., James R.,  
RA Kaur A., Madan A., Owen M.P., Ratcliffe A., Shaffer T., Hood L.;  
RT "Sequencing of human chromosome 14q24.3 region."  
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE OF 539-847 AND 1397-1822 FROM N.A.  
RC TISSUE=Brain;  
RX MEDLINE=95319502; PubMed=7596406;  
RA Sherrington R., Rogaev E.I., Liang Y., Rogaeva E.A., Levesque G.,  
RA Ikeda M., Chi H., Lin C., Li G., Holman K., Tsuda T., Mar L.,  
RA Foncin J.-F., Bruni A.C., Montesi M.P., Sorbi S., Rainero I.,  
RA Pinessi L., Nee L., Chumakov I., Pollen D., Brookes A.,  
RA Sanseau P., Polinsky R.J., Wasco W., da Silva H.A.R., Haines J.L.,  
RA Pericak-Vance M.A., Tanzi R.E., Roses A.D., Fraser P.E.,  
RA Rommens J.M., St George-Hyslop P.H.;  
RT "Cloning of a gene bearing missense mutations in early-onset familial  
RT Alzheimer's disease."  
RL Nature 375:754-760(1995).  
CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).

CC -!- CAUTION: Ref.2 sequence differs from that shown due to a  
CC frameshift in position 1661.

CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; AC007956; AAF61275.1; -.

DR EMBL; L40403; AAC42008.1; ALT\_FRAME.

DR EMBL; L40400; AAC42006.1; -.

KW Nuclear protein.

FT DOMAIN 15 205 PRO-RICH.

FT DOMAIN 382 430 GLN-RICH.

FT DOMAIN 807 1209 ARG-RICH.

FT DOMAIN 1488 1577 ARG-RICH.

FT CONFLICT 621 621 P -> S (IN REF. 2).

FT CONFLICT 1404 1404 T -> I (IN REF. 2).

FT CONFLICT 1821 1821 K -> E (IN REF. 2).

SQ SEQUENCE 1822 AA; 204947 MW; 8E6CB83FE540C7D2 CRC64;

Query Match 55.1%; Score 49; DB 1; Length 1822;

Best Local Similarity 57.1%; Pred. No. 77;

Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 2 HQPPQPLPPTVMFP 15

| || |||| : |

Db 80 HLPPPPLPPPPVMP 93

#### RESULT 11

##### CASB\_HUMAN

ID CASB\_HUMAN STANDARD; PRT; 226 AA.

AC P05814;

DT 01-NOV-1988 (Rel. 09, Created)

DT 01-AUG-1992 (Rel. 23, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Beta casein precursor.

GN CSN2 OR CASB.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Breast;

RA Menon R.S.;

RL Submitted (OCT-1989) to the EMBL/GenBank/DDBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=90353560; PubMed=2387396;

RA Loennerdal B., Bergstroem S., Andersson Y., Hjalmarsson K.,

RA Sundqvist A.K., Hernell O.;

RT "Cloning and sequencing of a cDNA encoding human milk beta-casein.";

RL FEBS Lett. 269:153-156(1990).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Placenta;  
 RX MEDLINE=94156198; PubMed=8112603;  
 RA Hansson L., Edlund A., Johansson T., Hernell O., Stroemqvist M.,  
 RA Lindquist S., Loennerdal B., Bergstroem S.;  
 RT "Structure of the human beta-casein encoding gene.";  
 RL Gene 139:193-199(1994).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Placenta;  
 RA Kwiatkowski D.J.;  
 RL Submitted (DEC-1997) to the EMBL/GenBank/DDBJ databases.  
 RN [5]  
 RP SEQUENCE OF 161-226 FROM N.A.  
 RC TISSUE=Breast;  
 RX MEDLINE=89240053; PubMed=2717418;  
 RA Menon R.S., Ham R.G.;  
 RT "Human beta-casein: partial cDNA sequence and apparent polymorphism.";  
 RL Nucleic Acids Res. 17:2869-2869(1989).  
 RN [6]  
 RP SEQUENCE OF 16-226.  
 RX MEDLINE=84185624; PubMed=6715339;  
 RA Greenberg R., Groves M.L., Dower H.J.;  
 RT "Human beta-casein. Amino acid sequence and identification of  
 phosphorylation sites.";  
 RL J. Biol. Chem. 259:5132-5138(1984).  
 CC -!- FUNCTION: Important role in determination of the surface  
 properties of the casein micelles.  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- TISSUE SPECIFICITY: Mammary gland specific. Secreted in milk.  
 CC -!- SIMILARITY: Belongs to the beta-casein family.  
 CC -!- DATABASE: NAME=Protein Spotlight;  
 CC NOTE=Issue 16 of November 2001;  
 CC WWW="http://www.expasy.org/spotlight/articles/sptlt016.html".  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X17070; CAA34916.1; -.  
 DR EMBL; X13766; CAA32017.1; -.  
 DR EMBL; AF027807; AAC82978.1; -.  
 DR EMBL; X55739; CAA39270.1; -.  
 DR EMBL; A24287; CAA01728.1; -.  
 DR EMBL; A30262; CAA02017.1; -.  
 DR PIR; I53730; KBHU.  
 DR Genew; HGNC:2447; CSN2.  
 DR MIM; 115460; -.  
 DR GO; GO:0005509; F:calcium ion binding; TAS.  
 DR GO; GO:0004857; F:enzyme inhibitor activity; TAS.  
 DR GO; GO:0006816; P:calcium ion transport; TAS.

DR InterPro; IPR001588; Casein.  
 DR Pfam; PF00363; caseins; 1.  
 DR PROSITE; PS00306; CASEIN\_ALPHA\_BETA; 1.  
 KW Milk; Phosphorylation; Glycoprotein; Signal.  
 FT SIGNAL 1 15  
 FT CHAIN 16 226 BETA CASEIN.  
 FT MOD\_RES 18 18 PHOSPHORYLATION.  
 FT MOD\_RES 21 21 PHOSPHORYLATION.  
 FT MOD\_RES 23 23 PHOSPHORYLATION.  
 FT MOD\_RES 24 24 PHOSPHORYLATION.  
 FT MOD\_RES 25 25 PHOSPHORYLATION.  
 FT CONFLICT 30 30 T -> P (IN REF. 6).  
 FT CONFLICT 34 34 MISSING (IN REF. 2).  
 FT CONFLICT 48 50 EDE -> TDQ (IN REF. 6).  
 FT CONFLICT 120 120 S -> Q (IN REF. 6).  
 FT CONFLICT 133 133 L -> V (IN REF. 1).  
 FT CONFLICT 140 140 H -> Q (IN REF. 1).  
 FT CONFLICT 149 149 L -> S (IN REF. 6).  
 FT CONFLICT 173 173 Q -> E (IN REF. 6).  
 FT CONFLICT 182 184 QVV -> EVL (IN REF. 6).  
 FT CONFLICT 188 188 Q -> V (IN REF. 6).  
 FT CONFLICT 207 207 T -> P (IN REF. 6).  
 FT CONFLICT 214 222 TQPLAPVHN -> PEPSTTZABH (IN REF. 6).  
 SQ SEQUENCE 226 AA; 25382 MW; 2619C524EA1358E8 CRC64;

Query Match 52.8%; Score 47; DB 1; Length 226;  
 Best Local Similarity 53.3%; Pred. No. 16;  
 Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 1 MHQPPQPLPPTVMFP 15  
 | | |||:| |: |  
 Db 150 MQQVPQPIQTLALP 164

# RESULT 12

## KNIR\_DROME

ID KNIR\_DROME STANDARD; PRT; 429 AA.  
 AC P10734; Q9VPC6;  
 DT 01-JUL-1989 (Rel. 11, Created)  
 DT 01-JUL-1989 (Rel. 11, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Zygotic gap protein knirps.  
 GN KNI OR NR0A1 OR CG4717.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Oregon-R; TISSUE=Salivary gland;  
 RX MEDLINE=89057148; PubMed=2904128;  
 RA Nauber U., Pankratz M.J., Kilnlin A., Seyffert E., Klemm U.,  
 RA Jackle H.;  
 RT "Abdominal segmentation of the Drosophila embryo requires a hormone  
 RT receptor-like protein encoded by the gap gene knirps."  
 RL Nature 336:489-492(1988).

RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkeley;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,  
 RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195(2000).  
 RN [3]  
 RP CHARACTERIZATION.  
 RX MEDLINE=96312963; PubMed=8670869;  
 RA Arnosti D.N., Gray S., Barolo S., Zhou J., Levine M.;  
 RT "The gap protein knirps mediates both quenching and direct repression  
 RT in the *Drosophila* embryo.";  
 RL EMBO J. 15:3659-3666(1996).  
 CC -!- FUNCTION: TRANSCRIPTIONAL REPRESSOR. BINDS TO MULTIPLE SITES IN  
 CC THE EVE STRIPE 3 ENHANCER ELEMENT. PLAYS AN ESSENTIAL ROLE IN THE  
 CC SEGMENTATION PROCESS BOTH BY REFINING THE EXPRESSION PATTERNS OF  
 CC GAP GENES AND BY ESTABLISHING PAIR-RULES STRIPES OF GENE  
 CC EXPRESSION.  
 CC -!- SUBCELLULAR LOCATION: Nuclear.  
 CC -!- SIMILARITY: Belongs to the nuclear hormone receptor family. NR0  
 CC subfamily.



CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X13331; CAA31709.1; -.  
 DR EMBL; AE003592; AAF51629.2; -.  
 DR PIR; S01919; S01919.  
 DR HSSP; P03372; 1HCP.  
 DR TRANSFAC; T00445; -.  
 DR FlyBase; FBgn0001320; kni.  
 DR GO; GO:0004879; F:ligand-dependent nuclear receptor activity; NAS.  
 DR GO; GO:0007088; P:regulation of mitosis; IMP.  
 DR InterPro; IPR001628; Znf\_C4steroid.  
 DR Pfam; PF00105; zf-C4; 1.  
 DR PRINTS; PR00047; STROIDFINGER.  
 DR ProDom; PD000035; Znf\_C4steroid; 1.  
 DR SMART; SM00399; ZnF\_C4; 1.  
 DR PROSITE; PS00031; NUCLEAR\_RECEPTOR; 1.  
 KW Receptor; Transcription regulation; DNA-binding; Nuclear protein;  
 KW Zinc-finger; Developmental protein; Repressor.  
 FT DNA\_BIND 5 71 NUCLEAR RECEPTOR-TYPE.  
 FT ZN\_FING 5 25 C4-TYPE.  
 FT ZN\_FING 42 66 C4-TYPE.  
 FT DOMAIN 97 101 POLY-ALA.  
 FT DOMAIN 137 142 POLY-HIS.  
 FT DOMAIN 143 149 POLY-GLN.  
 FT DOMAIN 200 213 POLY-ALA.  
 FT DOMAIN 375 382 POLY-SER.  
 SQ SEQUENCE 429 AA; 45611 MW; 79CEE86A66AB00C7 CRC64;

Query Match 52.8%; Score 47; DB 1; Length 429;  
 Best Local Similarity 57.1%; Pred. No. 32;  
 Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 2 HQPPQPLPPTVMFP 15  
 || | ||| :|||  
 Db 183 HQSPFQLPPLLFP 196

# RESULT 13

## ATIN\_HSV1F

ID ATIN\_HSV1F STANDARD; PRT; 479 AA.  
 AC P04486;  
 DT 13-AUG-1987 (Rel. 05, Created)  
 DT 13-AUG-1987 (Rel. 05, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE Alpha trans-inducing protein (VMW65) (ICP25) (VP16 protein)  
 DE (Alpha-TIF).  
 GN UL48.  
 OS Herpes simplex virus (type 1 / strain F).  
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Alphaherpesvirinae; Simplexvirus.

OX NCBI\_TaxID=10304;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=85298259; PubMed=2994050;  
 RA Pellett P.E., McKnight J.L.C., Jenkins F.J., Roizman B.;  
 RT "Nucleotide sequence and predicted amino acid sequence of a protein  
 RT encoded in a small herpes simplex virus DNA fragment capable of  
 RT trans-inducing alpha genes.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 82:5870-5874(1985).  
 CC -!- FUNCTION: RESPONSIBLE FOR TRANSCRIPTIONAL ACTIVATION OF IMMEDIATE  
 CC EARLY PROMOTERS (ALPHA GENES). ACTIVATION REQUIRES THE FORMATION  
 CC OF A HETEROMERIC COMPLEX WITH THE HOST CELL FACTOR. THESE TWO  
 CC PROTEINS THEN ASSEMBLE WITH THE OCTAMER MOTIF-BINDING PROTEIN  
 CC OCT-1 ON THE CIS-ACTING TARGET SEQUENCE: TAATGARAT.  
 CC -!- SIMILARITY: TO OTHER HERPESVIRUSES ALPHA TRANS-INDUCING PROTEIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; K03350; AAA45766.1; -.  
 DR PIR; A03727; IXBELF.  
 DR InterPro; IPR003174; Alpha\_TIF.  
 DR Pfam; PF02232; Alpha\_TIF; 1.  
 KW Transcription regulation; Trans-acting factor; DNA-binding.  
 SQ SEQUENCE 479 AA; 53053 MW; 8DFF24AC1717A1C6 CRC64;

Query Match 52.8%; Score 47; DB 1; Length 479;  
 Best Local Similarity 50.0%; Pred. No. 35;  
 Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 1 MHQPPQPLPPTVMF 14  
 | |||:| :|  
 Db 38 MPSPPMPVPPAALF 51

#### RESULT 14

ATIN\_HSV11

ID ATIN\_HSV11 STANDARD; PRT; 490 AA.  
 AC P06492;  
 DT 01-JAN-1988 (Rel. 06, Created)  
 DT 01-JAN-1988 (Rel. 06, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Alpha trans-inducing protein (VMW65) (ICP25) (VP16 protein)  
 DE (Alpha-TIF).  
 GN UL48.  
 OS Herpes simplex virus (type 1 / strain 17).  
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Alphaherpesvirinae; Simplexvirus.  
 OX NCBI\_TaxID=10299;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=88274327; PubMed=2839594;

RA McGeoch D.J., Dalrymple M.A., Davison A.J., Dolan A., Frame M.C.,  
 RA McNab D., Perry L.J., Scott J.E., Taylor P.;  
 RT "The complete DNA sequence of the long unique region in the genome of  
 RT herpes simplex virus type 1.";  
 RL J. Gen. Virol. 69:1531-1574(1988).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=86067203; PubMed=2999707;  
 RA Dalrymple M.A., McGeoch D.J., Davison A.J., Preston C.M.;  
 RT "DNA sequence of the herpes simplex virus type 1 gene whose product  
 RT is responsible for transcriptional activation of immediate early  
 RT promoters.";  
 RL Nucleic Acids Res. 13:7865-7879(1985).  
 RN [3]  
 RP DNA-BINDING.  
 RX MEDLINE=90005439; PubMed=2676518;  
 RA Cousens D.J., Greaves R., Goding C.R., O'Hare P.;  
 RT "The C-terminal 79 amino acids of the herpes simplex virus regulatory  
 RT protein, Vmw65, efficiently activate transcription in yeast and  
 RT mammalian cells in chimeric DNA-binding proteins.";  
 RL EMBO J. 8:2337-2342(1989).  
 CC -!- FUNCTION: RESPONSIBLE FOR TRANSCRIPTIONAL ACTIVATION OF IMMEDIATE  
 CC EARLY PROMOTERS (ALPHA GENES). ACTIVATION REQUIRES THE FORMATION  
 CC OF A HETEROMERIC COMPLEX WITH THE HOST CELL FACTOR. THESE TWO  
 CC PROTEINS THEN ASSEMBLE WITH THE OCTAMER MOTIF-BINDING PROTEIN  
 CC OCT-1 ON THE CIS-ACTING TARGET SEQUENCE: TAATGARAT.  
 CC -!- SIMILARITY: TO OTHER HERPESVIRUSES ALPHA TRANS-INDUCING PROTEIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X14112; CAA32298.1; -.  
 DR EMBL; X03141; CAA26913.1; -.  
 DR PIR; A24118; IXBE17.  
 DR PDB; 16VP; 26-SEP-01.  
 DR TRANSFAC; T00894; -.  
 DR InterPro; IPR003174; Alpha\_TIF.  
 DR Pfam; PF02232; Alpha\_TIF; 1.  
 KW Transcription regulation; Trans-acting factor; DNA-binding;  
 KW 3D-structure.  
 FT DNA\_BIND 411 490 EXPERIMENTALLY DEDUCED.  
 FT SITE 442 442 CRITICAL ROLE IN ACTIVATION.  
 SQ SEQUENCE 490 AA; 54345 MW; 8DDDEDEDB2A699D3 CRC64;

Query Match 52.8%; Score 47; DB 1; Length 490;  
 Best Local Similarity 50.0%; Pred. No. 36;  
 Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 1 MHQPPQPLPPTVMF 14  
 | |||:| :|  
 Db 49 MPSPPMPVPPAALF 62

## RESULT 15

ATIN\_HSV2H

ID ATIN\_HSV2H STANDARD; PRT; 490 AA.  
AC P23990; P29793;  
DT 01-APR-1993 (Rel. 25, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE Alpha trans-inducing protein (VMW65) (ICP25) (VP16 protein)  
DE (Alpha-TIF).  
GN VP16 OR UL48.  
OS Herpes simplex virus (type 2 / strain HG52), and  
OS Herpes simplex virus (type 2 / strain 333).  
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
OC Alphaherpesvirinae; Simplexvirus.  
OX NCBI\_TaxID=10315, 10313;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=HG52;  
RX MEDLINE=91365250; PubMed=1653757;  
RA Cress A., Triezenberg S.J.;  
RT "Nucleotide and deduced amino acid sequences of the gene encoding  
RT virion protein 16 of herpes simplex virus type 2.";  
RL Gene 103:235-238(1991).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=HG52;  
RA Dolan A.;  
RL Submitted (FEB-1997) to the EMBL/GenBank/DDBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=333;  
RX MEDLINE=92046332; PubMed=1658370;  
RA Greaves R.F., O'Hare P.;  
RT "Sequence, function, and regulation of the Vmw65 gene of herpes  
RT simplex virus type 2.";  
RL J. Virol. 65:6705-6713(1991).  
CC -!- FUNCTION: RESPONSIBLE FOR TRANSCRIPTIONAL ACTIVATION OF IMMEDIATE  
CC EARLY PROMOTERS (ALPHA GENES).  
CC -!- SIMILARITY: TO OTHER HERPESVIRUSES ALPHA TRANS-INDUCING PROTEIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M60050; AAA45863.1; -.  
DR EMBL; Z86099; CAB06734.1; -.  
DR EMBL; M75098; AAA45862.1; -.  
DR InterPro; IPR003174; Alpha\_TIF.  
DR Pfam; PF02232; Alpha\_TIF; 1.  
KW Transcription regulation; Trans-acting factor; DNA-binding.  
FT SITE 443 443 CRITICAL ROLE IN ACTIVATION.  
FT CONFLICT 12 12 A -> R (IN REF. 1).

SQ SEQUENCE 490 AA; 54620 MW; 2E9FDA8D0D8BC174 CRC64;

Query Match 52.8%; Score 47; DB 1; Length 490;

Best Local Similarity 50.0%; Pred. No. 36;

Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 1 MHQPPQPLPPTVMF 14

| || |:|| :|

Db 47 MPSPPMPVPPAALF 60

Search completed: August 24, 2004, 15:43:47

Job time : 10.0597 secs